



UNIVERSITAT DE
BARCELONA

DEPARTAMENTO DE FILOLOGÍA HISPÁNICA, TEORÍA DE LA LITERATURA Y COMUNICACIÓN.

MÁSTER DE ESPAÑOL COMO LENGUA EXTRANJERA EN ÁMBITOS PROFESIONALES 2018-2020

**Propuesta de aplicación didáctica de los corpus paralelos en el aula
de ELE para los estudiantes sinohablantes**

Weiqian Wei

Trabajo final de Máster

Dirección:

Dra. Mar Cruz Piñol

Universidad de Barcelona

Mar
C. P.

Firmado digitalmente
por Mar C. P.
Nombre de
reconocimiento (DN):
cn=Mar C. P.,
o=Universitat de
Barcelona, ou=Facultat
de Filologia,
email=mcruz@ub.edu,
c=ES
Fecha: 2020.08.27
11:04:01 +01'00'

AGRADECIMIENTOS

A mis padres, por su amor y apoyo incondicional.

A mis amigos, Chenxi, Shiyu y Weiken, por ser mi fuente de inspiración y energía en todo momento para realizar este trabajo.

A la profesora Mar Cruz Piñol, por haber sido mi guía y mi estímulo durante estos meses que nos han unido. Gracias por haberme orientado a lo largo de este camino, por compartir conmigo su sabiduría, por animarme cuando estaba perdida.

RESUMEN

En los últimos años, la Lingüística de corpus está suscitando un interés creciente entre el profesorado de lenguas extranjeras. Los corpus se consideran como un recurso productivo que puede aportar muchas ventajas y posibilidades en la enseñanza de lenguas. Un corpus puede incluir más de una lengua y, entre los corpus multilingües, se encuentran los *corpus paralelos*, que se están empleando cada vez más en el ámbito de la enseñanza de lenguas extranjeras. Con todo, hay que decir que un gran número de los estudios realizados en este ámbito se basan en los corpus paralelos que incluyen inglés y son exclusivamente relacionados con la enseñanza de inglés, mientras que la aplicación de los corpus paralelos dirigida a los aprendices sinohablantes de español ha sido relativamente poco estudiada. Conscientes de ello, este trabajo tiene como objetivos presentar las posibles aplicaciones didácticas de los corpus paralelos, comparar los corpus paralelos chino-español disponibles en línea y lanzar unas ideas sobre el uso directo de este tipo de recurso en el aula de español como lengua extranjera con especial atención a los estudiantes sinohablantes.

Palabras clave: corpus, corpus paralelo, uso directo, enseñanza de lenguas extranjeras, español como lengua extranjera, estudiantes sinohablantes

ABSTRACT

In recent years, Corpus Linguistics has been attracting increasing interest among foreign language teachers. Corpus has been considered as a productive resource that can bring many advantages and possibilities in language teaching. A corpus does not necessarily contain only one language. The parallel corpus is a type of multilingual corpus which is increasingly being used in the field of foreign language teaching. However, a large number of studies carried out in this field are based on parallel corpus that include English and are exclusively related to the teaching of English. The application of parallel corpus aimed at Chinese-speaking learners of Spanish has been relatively unexplored. Aware of this, this work aims to present the possible didactic applications of parallel corpus, compare the Chinese-Spanish parallel corpus available online and propose some practical examples of the direct use of this type of resource in the classroom of Spanish as a foreign language, with special attention to Chinese-speaking students.

Key words: corpus, parallel corpus, direct use, language teaching, Spanish as a foreign language, Chinese-speaking learners

ÍNDICE

1. INTRODUCCIÓN	9
2. OBJETIVOS	11
3. MARCO TEÓRICO.....	12
3.1 Lingüística de corpus	12
3.1.1 Definición y características.....	12
3.1.2 Conceptos clave	13
3.1.2.1 Concordancias.....	13
3.1.2.2 Listados de frecuencias.....	17
3.1.2.3 Lematización.....	18
3.1.3 Tipología de corpus.....	19
3.1.4 Aplicaciones didácticas de los corpus en el ámbito del ELE.....	23
3.2 Corpus paralelos y su uso práctico en el ELE.....	29
3.2.1 Definición y tipología	29
3.2.2 Diseño y elaboración.....	31
3.2.3 La web como corpus paralelo	37
3.2.4 Aplicaciones didácticas de los corpus paralelos	38
3.2.4.1 Uso directo en el aula de la consulta de concordancias paralelas.....	40
3.2.4.2 Uso indirecto de los corpus paralelos para la elaboración de materiales didácticos	44
3.2.4.3 Uso de los corpus paralelos como herramienta de consulta para mejorar la expresión escrita	45
3.2.4.4 Uso de los corpus paralelos como base para la investigación en lingüística contrastiva	46
4. METODOLOGÍA	49
4.1 Justificación de selección de los corpus.....	49

4.2 Criterios de análisis.....	53
5. PRESENTACIÓN Y ANÁLISIS DE LOS CORPUS SELECCIONADOS.....	56
5.1 TAUS Data Cloud	56
5.2 Glosbe	58
5.3 Xiyuzhushou (Eshelper).....	61
5.4 ReversoContext.....	64
5.5 CPEIC	67
5.6 Análisis de los corpus paralelos chino-español seleccionados	71
6. PROPUESTA DE ACTIVIDADES	75
6.1 Introducción	75
6.2 Actividades para aprender locuciones.....	76
6.3 Actividades para aprender marcadores del discurso	86
6.4 Otras posibilidades.....	92
7. CONCLUSIONES Y LÍNEAS FUTURAS DE INVESTIGACIÓN.....	94
8. BIBLIOGRAFÍA	97
9. ANEXOS	108

ÍNDICE DE ILUSTRACIONES, TABLAS Y FIGURAS

Ilustración 1 Consulta en <i>CREA</i>	14
Ilustración 2 Consulta en <i>OPUS</i>	15
Ilustración 3 Consulta en <i>CPEIC</i>	16
Ilustración 4 Listado de las 1000 palabras más frecuentes en <i>CREA</i>	17
Ilustración 5 Listado de concordancias del verbo <i>hablar</i> presente en el corpus paralelo <i>Eshelper</i>	18
Ilustración 6 Listado de concordancias del verbo <i>hablar</i> presente en el corpus paralelo <i>MyMemory</i>	19
Ilustración 7 Consulta en <i>Linguee</i>	22
Ilustración 8 Muestra de corpus <i>PaChEs</i>	34
Ilustración 9 Consulta en <i>RST Spanish-Chinese Treebank</i> (1)	52
Ilustración 10 Consulta en <i>RST Spanish-Chinese Treebank</i> (2)	53
Ilustración 11 Búsqueda de la combinación léxica <i>me gusta</i> en <i>TAUS Data Cloud</i> ...	57
Ilustración 12 Búsqueda de la palabra <i>gustar</i> en <i>Glosbe</i> (1)	59
Ilustración 13 Búsqueda de la palabra <i>gustar</i> en <i>Glosbe</i> (2)	59
Ilustración 14 Búsqueda de la palabra <i>gustar</i> en <i>Eshelper</i> (1)	62
Ilustración 15 Búsqueda de la palabra <i>gustar</i> en <i>Eshelper</i> (2)	63
Ilustración 16 Búsqueda de la palabra <i>gustar</i> en <i>ReversoContext</i>	65
Ilustración 17 Búsqueda de la palabra <i>gustar</i> en <i>CPEIC</i>	67
Ilustración 18 Búsqueda de la palabra <i>hablar</i> en <i>Glosbe</i>	71
Ilustración 19 Búsqueda de la palabra <i>mano</i> en <i>Eshelper</i>	84
Ilustración 20 Búsqueda de la combinación léxica <i>de primera mano</i> en <i>ReversoContext</i>	85
Ilustración 21 Búsqueda de la palabra <i>aunque</i> en <i>ReversoContext</i>	88
Ilustración 22 Búsqueda de la palabra <i>aunque</i> en <i>Eshelper</i>	90

Tabla 1 Comparación de Taus Data Cloud, Glosbe, ReversoContext, Eshelper y CPEIC. Elaboración propia.....	70
Tabla 2 Traducción de <i>aunque</i> en 20 ejemplos. Elaboración propia.	91
Figura 1 Tipología de los corpus bi-/multilingües, elaboración propia a partir del esquema que propone Doval (2017)	31

ÍNDICE DE ABREVIATURAS

CREA	Corpus de Referencia del Español Actual
DLE	Diccionario de la Lengua Española
EFE	Español con Fines Específicos
ELE	Español como Lengua Extranjera
ELEAP	Español como Lengua Extranjera en Ámbitos Profesionales
KWIC	Key Word in Context
LC	Lingüística de corpus
TM	Memoria de traducción

1. INTRODUCCIÓN

Los corpus, considerados como un recurso que nos ofrece abundantes muestras reales de una lengua, están despertando un interés creciente entre el profesorado de lenguas extranjeras y se están empleando cada día más en el ámbito de Español como Lengua Extranjera (ELE). Entre los corpus multilingües, se encuentran los corpus paralelos, que se caracterizan por contener la misma colección de textos en dos o más idiomas, uno de ellos es el idioma original y los otros la traducción (McEnery y Hardie, 2012). En comparación con los corpus monolingües, los corpus paralelos aportan las ventajas como por ejemplo mostrar la correspondencia y la contrastividad entre dos lenguas (Barlow, 2000; Kenning, 2010). Los corpus paralelos juegan un papel destacable especialmente para las investigaciones en traducción, los estudios contrastivos y la lingüística computacional. Al mismo tiempo, se están empleando cada día más en el ámbito de ELE. Sin embargo, comparando con la enseñanza de inglés, el empleo de este recurso en el ámbito de la enseñanza del ELE para los estudiantes sinohablantes ha sido relativamente poco estudiado.

Con respecto a mi interés personal para elegir este tema, mi experiencia al hacer las prácticas del máster ELEAP (Español como Lengua Extranjera en Ámbitos Profesionales) en el proyecto PaChEs¹ me impulsó a realizar esta investigación. Gracias a la valiosa oportunidad ofrecida por la Universitat de Barcelona, los alumnos del mencionado máster, pudimos participar en este proyecto para construir un corpus paralelo que puede servir para innumerables investigaciones en el futuro. Durante el proceso de hacer las prácticas, fuimos observando con especial atención la

¹ Un corpus paralelo chino-español en construcción actualmente. El proyecto está a cargo de un grupo de investigación formado por los profesores y estudiantes de la Universidad de Santiago de Compostela, la Universidad de Barcelona y la Fundación de Instituto Confucio de Barcelona.

contrastividad entre chino y español. De ahí que se despertara en mí el interés por investigar la aplicación didáctica de este tipo de recurso.

Por todo lo expuesto, los objetivos que se pretende conseguir con este trabajo (véase el apartado 2) consisten en conocer las aplicaciones didácticas de los corpus paralelos y seleccionar unos corpus paralelos chino-español disponibles en línea para lanzar unas ideas sobre la utilización de los corpus paralelos en la enseñanza del ELE, con especial atención a los estudiantes sinohablantes. Conscientes de ello, el presente trabajo consta de seis apartados fundamentales:

En este primer y el siguiente apartado, hacemos una breve introducción sobre nuestro trabajo y presentamos las preguntas de investigación y los objetivos que pretendemos alcanzar. En el apartado 3, nos centraremos en el marco teórico, que está compuesto por dos partes principales: por una parte, la Lingüística de corpus (LC) en general y, por otra parte, los corpus paralelos y sus usos prácticos en ELE. En el apartado 4, explicaremos la metodología adoptada para seleccionar los corpus paralelos adecuados para utilizar en el aula de ELE. Después, en el apartado 5 presentamos los corpus paralelos chino-español disponibles en línea que hemos seleccionado y describiremos las características de cada uno. Luego, en el apartado 6 propondremos unas actividades y posibles usos de los corpus paralelos en la enseñanza del ELE para los estudiantes sinohablantes. Por último, en el apartado 7 se mostrarán las conclusiones principales y las líneas futuras de investigación que se derivan de nuestro trabajo.

2. OBJETIVOS

En el presente trabajo, se trata de responder las siguientes dos preguntas:

- a) ¿Qué posibilidades pueden aportar los corpus paralelos en la enseñanza del ELE, especialmente para los alumnos sinohablantes?
- b) En cuanto a los corpus paralelos chino-español disponibles en línea actualmente, ¿cuáles son los adecuados para emplear en el aula de ELE?

Asimismo, los objetivos que se ha pretendido conseguir con este trabajo fin de máster son los siguientes:

- Objetivo 1: Conocer qué es un corpus paralelo y cuáles son las aplicaciones didácticas que pueden tener en la enseñanza del ELE, con especial atención a los estudiantes sinohablantes.
- Objetivo 2: Comparar los corpus paralelos chino-español disponibles en línea actualmente y seleccionar los adecuados para la elaboración de la propuesta didáctica.
- Objetivo 3: Presentar unas propuestas de actividades utilizando los corpus paralelos chino-español para los alumnos sinohablantes.

3. MARCO TEÓRICO

3.1 Lingüística de corpus

3.1.1 Definición y características

Para poder entender los corpus paralelos y usarlos en la enseñanza de lenguas, parece fundamental saber qué es un corpus y para qué sirve. De ahí que se exija una descripción previa sobre la disciplina “Lingüística de Corpus” (LC) y los conceptos clave en este ámbito. A la hora de investigar en el ámbito de la LC, la primera pregunta que se debe plantear es la de encontrar una definición clara de *corpus lingüístico*. Mientras que se ha sugerido una gran variedad de definiciones para el término *corpus*, en este trabajo fin de máster utilizamos la definición propuesta por Sinclair (2004), quien define *corpus* como sigue:

A corpus is a collection of pieces of language text in electronic form, selected according to external criteria to represent, as far as possible, a language or language variety as a source data for linguistic research (Sinclair, 2004, cap. 1²).

Esta definición nos permite extraer las características principales del corpus. En otras palabras, un *corpus* debe cumplir los siguientes requisitos:

- 1) Formato electrónico: se trata de una recopilación de textos de varios millones de palabras que, actualmente, debe estar informatizado. Es decir, los textos incluidos deben estar en formato electrónico, y el corpus debe permitir hacer diferentes búsquedas entre grandes cantidades de texto.
- 2) Selección de textos: la selección de los textos debe seguir ciertos criterios objetivos para que sean representativos y muestren la realidad de la lengua meta o una variedad de una lengua. Por ello, los investigadores pueden recoger todas las

² http://icar.cnrs.fr/ecole_thematique/contaci/documents/Baude/wynne.pdf

estructuras posibles en la lengua meta siguiendo los criterios objetivos. Cabe destacar aquí la importancia de la representatividad en el proceso de la selección de textos y compilación de corpus, puesto que la representatividad permite que sus usuarios obtengan resultados y conclusiones fiables desde los datos estadísticos (Cruz, 2017).

- 3) Por último, el corpus debe ser una fuente de datos para investigaciones lingüísticas.

3.1.2 Conceptos clave

Conocer los conceptos clave en la LC es de suma importancia para nuestro trabajo a la hora de comparar los corpus paralelos chino-español disponibles en línea actualmente, y establecer una serie de criterios para seleccionar los adecuados para utilizar en el aula de ELE (objetivo 2). En este apartado, por lo tanto, vamos a definir los conceptos de concordancias, listados de frecuencias, y lematización.

3.1.2.1 Concordancias

Las concordancias son una herramienta ampliamente usada por los investigadores y lingüistas que trabajan con los corpus. Pérez Hernández (2002) nos presenta la definición de concordancia como sigue:

Una concordancia, normalmente llamada *KWIC* (*Key Word in Context*) es una colección que recoge todas las apariciones de una palabra en un texto o conjunto de textos, junto con un número determinado [...] de caracteres de co-texto anterior y posterior (la palabra que se está estudiando o no, suele aparecer en medio, resaltada en pantalla con un formato o color diferente). De esta forma es posible visualizar a la vez una gran cantidad de ejemplos de uso de una palabra o un grupo de palabras (Pérez Hernández, 2002)

Como ejemplo, presentamos las concordancias que obtenemos a través de la búsqueda de la palabra *ejemplo* en el corpus CREA³ (como se obtienen demasiados resultados, la búsqueda ha sido filtrada):

Concordancias (RAE)

Consulta: **ejemplo, en Libros, en CREA, en ESPAÑA**
 Resultado: **(Filtrado) 276 casos en 9 documentos.**

OBTENCIÓN DE EJEMPLOS

Recuperar Concordancias: Normal Clasificación: Agrupación: Marcas:

Cómo citar el CORPUS Concordancias.

Pantalla: 1 de 12. Siguiente 1 2 3 4 5 6 7 8 9 10 11 12 Ver párrafos

Nº	CONCORDANCIA	AÑO	AUTOR	TÍTULO
1	, instalación limpia) del programa. Si desea, por ejemplo, mantener una copia de su antiguo sistema ope	** 2000	Pardo Niebla, Miguel	Windows
2	abitualmente nuestras tareas cotidianas. Así, por ejemplo, en Windows 2000, los documentos se organizan	** 2000	Pardo Niebla, Miguel	Windows
3	e que describen su contenido o utilidad. Así por ejemplo, un programa de tratamiento de textos como Wo	** 2000	Pardo Niebla, Miguel	Windows
4	n gráfica en la interfaz de Windows 2000. Así por ejemplo, el ratón, está representado en la pantalla d	** 2000	Pardo Niebla, Miguel	Windows
5	como carpetas, unidades de disco o impresoras. Un ejemplo claro, es el icono Mi PC que aparece cuando i	** 2000	Pardo Niebla, Miguel	Windows
6	ntenido de los elementos que representan. Así por ejemplo, haciendo doble clic sobre el icono de una un	** 2000	Pardo Niebla, Miguel	Windows
7	os elementos contenidos en el escritorio. Así por ejemplo, para seleccionar cualquiera de los iconos de	** 2000	Pardo Niebla, Miguel	Windows
8	na que contiene nuevos iconos o carpetas como por ejemplo, Mi PC. Finalmente, también es posible dispon	** 2000	Pardo Niebla, Miguel	Windows
9	l escritorio o sobre algún otro elemento como por ejemplo la Papelera de reciclaje, el icono desaparece	** 2000	Pardo Niebla, Miguel	Windows
10	e determinados iconos o tipos de iconos, como por ejemplo Vaciar Papelera de reciclaje (para eliminar e	** 2000	Pardo Niebla, Miguel	Windows
11	nicio 2.3.5. El menú Inicio El menú Inicio es un ejemplo representativo del funcionamiento de los menús	** 2000	Pardo Niebla, Miguel	Windows
12	asado pequeña para nuestras necesidades. Así por ejemplo, si deseamos trabajar con una aplicación maxi	** 2000	Pardo Niebla, Miguel	Windows
13	do cualquier mapa de bits para el fondo (como por ejemplo una fotografía digitalizada). El escritorio e	** 2000	Pardo Niebla, Miguel	Windows
14	r dos herramientas en paralelo. Si se piensa, por ejemplo, en alguien que se encuentre redactando un do	** 2000	Pardo Niebla, Miguel	Windows
15	de los laterales de la ventana. Si se pulsa, por ejemplo, Flecha decha. el cursor se situará en el late	** 2000	Pardo Niebla, Miguel	Windows

Ilustración 1 Consulta en CREA

Con este ejemplo podemos ver que las concordancias nos proporcionan la palabra que buscamos (*keyword*) en su contexto y, como indica Pérez Hernández (2002), la palabra *ejemplo* aparece en medio de cada ejemplo en diferente color. Cabe adelantar que, en nuestro caso, las concordancias en los corpus paralelos se verán de forma paralela en dos o varias lenguas. Por ejemplo, consultando en el corpus OPUS⁴ la palabra *ejemplo*, las concordancias se muestran de la siguiente forma:

³ Corpus de Referencia del Español Actual, disponible en: <http://www.rae.es>

⁴ Una colección de corpus paralelos multilingües disponibles en línea, disponible en <http://opus.nlpl.eu/bin/opuscqp.pl?corpus=MultiUN;lang=es>

OPUS – Corpus query (CWB)

corpora

MultiUN

languages

ar de en
es fr ru zh

CQP query (CWB)

show attributes

alignments

A CQP query consists of a regular expression over *attribute expressions*.
Introduction of the query syntax
Example queries

cooperación

☒ word
☐ deprel
☐ head
☐ lem
☐ svmtool
☐ tree

select

show max

20

hits

☐ vertical
☒ KWIC
☐ horizontal

(advanced search)

☐ ar
☐ de
☐ en
☐ fr
☐ ru
☒ zh

Query string: "decir" :ZH []

13 hits found

227323073	en inglés) : Sr . Presidente : Ante todo , deseo decir cuánto me place participar en esta sesión del Con
zh	川口女士 (日本) (以英语发言) : 主席先生 , 首先我要说 , 我多么高兴地参加今天在你主持下召开的安理会本次会议 。
227337551	s de cooperación bilateral . Esto es lo que quería decir . Junto con sus asociados de la Unión Europea , F
zh	这些是我要说的几点意见 。
307286916	lativa de convocar este importante debate . Huelga decir que las Naciones Unidas , y en particular el Cons
zh	不用说 , 联合国 、 特别是安全理事会在维持国际和平与安全中起着极为重要的作用 。
307297996	l Consejo de Seguridad . Sin embargo , está de más decir que la cooperación entre las Naciones Unidas y la
zh	然而 , 不用说 , 联合国与区域组织之间的合作是最重要的因素之一 。
307298482	na mayor cooperación entre esas entidades . Huelga decir que la cooperación entre el Consejo de Seguridad
zh	不用说 , 安全理事会与区域组织之间的合作 , 以及共同努力的协调将对实现安全及区域和国际稳定产生显著的影响 。
379419867	rrafo 20 , segunda frase Donde dice 30 países debe decir 35 países . En la nota 21 debe decir Afganistán ,
zh	将 “ 有13个国家 ” 改为 “ 有10个国家 ” 。

Ilustración 2 Consulta en *OPUS*

No obstante, en nuestro caso, conviene adelantar que la mayoría de los corpus paralelos chino-español disponibles en línea (como se verá en los apartados 4 y 5) que vamos a utilizar para la elaboración de la propuesta didáctica en este trabajo (objetivo 3) no presentan los resultados en forma de concordancias *KWIC*, es decir, las concordancias no están presentadas tan ordenadas y alineadas con el mismo número de palabras o caracteres en ambos lados. A diferencia de las concordancias *KWIC*, las concordancias paralelas⁵ se refieren a las listas de ocurrencias de una palabra o una combinación de palabra junto con sus contextos y alineadas con sus traducciones en una o más lenguas de forma paralela (Frankenberg-Garcia, 2005). Por ejemplo, la Ilustración 3 nos presenta el resultado de la búsqueda de la palabra *ejemplo* en el CPEIC (Corpus Paralelo de Español, Inglés y Chino)⁶, y las concordancias paralelas se muestran de la siguiente forma:

⁵ También denominadas las concordancias bilingües.

⁶ Corpus Paralelo de Español, Inglés y Chino (CPEIC), un corpus paralelo trilingüe elaborado por un grupo de la Universidad Nacional de Cheng Kung (NCKU) desde el año 2007 con el propósito de facilitar el análisis contrastivo entre las tres lenguas. Disponible en: <http://140.116.245.146/cpeic/TriLin.html?role=teacher>

Ilustración 3 Consulta en CPEIC

Con ayuda de las concordancias podemos obtener mucha información sobre dicho término en la lengua real. Las concordancias que ofrecen los corpus son una herramienta de suma importancia y muestran múltiples utilidades que pueden ser de gran ayuda a todas las personas interesadas en la didáctica de lenguas (Alonso Pérez-Ávila, 2006). Obtener las concordancias permitirá que el aprendiz de segunda lengua pueda observar cómo se utiliza esa palabra en la lengua real y en contexto. Tras observar los casos en un contexto real de manera repetida, el aprendiz tendrá acceso a extraer reglas gramaticales de modo inductivo, buscar las combinaciones léxicas, analizar los matices léxico-semánticos, saber en qué tipo de texto se utiliza la palabra buscada, etc. Teniendo en cuenta todas esas posibilidades se muestra que, en el momento de su aplicación a la didáctica de lenguas extranjeras, el uso de las concordancias es muy variado: se pueden utilizar para trabajar con contenidos léxicos, gramaticales, socioculturales, etc. En el apartado 3.1.4 vamos a hacer una descripción más detallada sobre la utilización de las concordancias en la didáctica de lenguas extranjeras.

3.1.2.2 Listados de frecuencias

Disponer de la información sobre las palabras más frecuentes de una lengua puede constituir una gran ventaja a la hora de elaborar materiales, diccionarios y otros libros de referencia (Cruz, 2017). Asimismo, las listas de frecuencias pueden ayudar a los alumnos a elegir las estructuras lingüísticas más habituales en el uso real de la lengua en su producción (Martos Eliche y Contreras Izquierdo, 2018). Seguidamente, en la Ilustración 4 podemos ver que el Corpus de Referencia del Español Actual (CREA) nos proporciona el listado de las 1000 palabras más frecuentes en lengua española:

Orden	Frec.absoluta	Frec.normalizada			
1. de	9,999,518	65545.55	21. lo	866,955	5682.77
2. la	6,277,560	41148.59	22. como	773,465	5069.96
3. que	4,681,839	30688.85	23. más	661,696	4337.33
4. el	4,569,652	29953.48	24. o	542,284	3554.60
5. en	4,234,281	27755.16	25. pero	450,512	2953.04
6. y	4,180,279	27401.19	26. sus	449,870	2948.84
7. a	3,260,939	21375.03	27. le	413,241	2708.74
8. los	2,618,657	17164.95	28. ha	380,339	2493.07
9. se	2,022,514	13257.31	29. me	374,368	2453.93
10. del	1,857,225	12173.87	30. si	327,480	2146.58
11. las	1,686,741	11056.37	31. sin	298,383	1955.86
12. un	1,659,827	10879.95	32. sobre	289,704	1898.97
13. por	1,561,904	10238.07	33. este	285,461	1871.16
14. con	1,481,607	9711.74	34. ya	274,177	1797.19
15. no	1,465,503	9606.18	35. entre	267,493	1753.38
16. una	1,347,603	8833.36	36. cuando	257,272	1686.38
17. su	1,103,617	7234.06	37. todo	247,340	1621.28
18. para	1,062,152	6962.26	38. esta	238,841	1565.57
19. es	1,019,669	6683.79	39. ser	232,924	1526.78
20. al	951,054	6234.03	40. son	232,415	1523.45

Ilustración 4 Listado de las 1000 palabras más frecuentes en *CREA*

En realidad, la mayoría de programas que procesan un corpus nos permite obtener la información de las palabras más frecuentes. Actualmente, este tipo de información se puede conseguir a partir de los corpus textuales informatizados, y de los listados y diccionarios de frecuencias electrónicos o impresos (Cruz, 2017).

3.1.2.3 Lematización

La lematización de un corpus consiste en vincular cada *forma* con su *lema*. La *forma* se refiere a la palabra como aparece en su contexto de uso, mientras que el *lema* es la entrada que permite la búsqueda en un diccionario. Tomando el caso del paradigma verbal como ejemplo, si buscamos el verbo en infinitivo (*hablar*) en el corpus paralelo chino-español lematizado *Eshelper* (Ilustración 5), obtendremos un listado de las concordancias de todas las formas de ese verbo encontradas en el corpus (*hablo*, *hablaste*, *hablando*...). Es decir, en este caso el lema es el verbo en infinitivo mientras que la forma consiste en los verbos flexionados.

hablar 专八 A1A2

音标: [aβ'lar] 发音: 全球发音: 生词本: 添加笔记:

西汉-汉西词典 西语动词变位 西语例句库 西语百科 中文百科 近反义词 西语词典 西语维基词典 西英词典 全文检索

有奖纠错 | 划词

一 原声例句

1. Sí, soy yo .¿Con quién **hablo**?
对的, 是我。您是哪位?
《商务西班牙语900句》
2. "Mañana **hablaré** con el conserje", pensé.
"明天我得和管理员谈一谈。"我这样想。
《新西语侦探mini小说》
3. **Hablar** (to speak/ talk) / **hablando** / **hablado**
vt. 说话, 讲话
《西语实用词句合集》
4. ¿Oigan?, acabo de **hablar** con el gerente...
听着, 我刚刚和经理谈了一下。
《咱们裸熊》

Ilustración 5 Listado de concordancias del verbo *hablar* presente en el corpus paralelo *Eshelper*⁷

Por el contrario, el propio corpus CREA no está lematizado, y de ahí que en la Ilustración 4 (ver página anterior) muestre por separado la frecuencia del artículo en singular y en plural, en masculino y en femenino. La ausencia de lematización muestra varias limitaciones al usar los corpus, especialmente cuando se trata de lenguas con

⁷ Disponible en: <https://www.esdict.cn/dicts/es/hablar>

mucha flexión, como en nuestro caso, el español. Si por ejemplo buscamos la misma palabra *hablar* en un corpus no lematizado como *MyMemory*⁸, solo obtendremos las concordancias que incluyen la forma *hablar* (Ilustración 6). Para localizar todas las formas de paradigma verbal habrá que introducir, una por una, todas las posibilidades de flexión.



The screenshot shows the MyMemory website interface. At the top, there is a logo for MyMemory (by translated LABS) and navigation links: "0 contribution(s)", "☆☆☆", "Translation API", "About MyMemory", and a user icon. Below the header is a table with three rows of concordances for the verb "hablar". Each row contains a Spanish sentence, a Chinese translation, and metadata including the last update date, usage frequency, quality, and a reference link to "MatteoT".

No quiero hablar más del tema.	我不想再进一步追溯。	Last Update: 2017-01-04 Usage Frequency: 1 Quality: ☆☆☆☆☆ Be the first to vote Reference: MatteoT
No fue una equivocación al hablar .	这不是疏忽，也不是失言。	Last Update: 2017-01-03 Usage Frequency: 1 Quality: ☆☆☆☆☆ Be the first to vote Reference: MatteoT
Discúlpeme por hablar en inglés.	(以英语发言)	Last Update: 2017-01-03 Usage Frequency: 1 Quality: ☆☆☆☆☆ Be the first to vote Reference: MatteoT

Ilustración 6 Listado de concordancias del verbo *hablar* presente en el corpus paralelo *MyMemory*

3.1.3 Tipología de corpus

Conocer la tipología de corpus sirve para que los profesores de lenguas extranjeras o segundas lenguas puedan decidir por sí mismos qué tipo de corpus les resultará más útil en cada situación (Cruz, 2017). Vargas Sierra (2006) y Cruz (2017) resumen la tipología de los corpus a partir de los seis criterios: los idiomas incluidos, el canal de producción, el grado de representatividad, la especificidad de los textos, el porcentaje y la distribución de los textos y el nivel de análisis. Pero de estos nos interesa especialmente la clasificación según los idiomas incluidos. Siguiendo este criterio, se distingue entre:

⁸ Disponible en: <https://mymemory.translated.net/en/Spanish/Chinese-%28Simplified%29/hablar>

- i. Corpus monolingües: los corpus que recogen muestras de textos en una única lengua.
- ii. Corpus bilingües o multilingües: los corpus que incluyen textos en dos o más lenguas, sin necesidad de traducciones.
- iii. Corpus comparables: los textos de este tipo de corpus son parecidos en cuanto a su contenido, en diferentes lenguas o diferentes variedades de una misma lengua.
- iv. Corpus paralelos: un corpus es paralelo cuando “contiene textos redactados en una lengua (la original) junto con las traducciones de éstos a otras lenguas (Vargas Sierra, 2006: 14)”. Los corpus paralelos, de los que realizaremos una descripción más detallada en el apartado 3.2, por su parte, como indica Vargas Sierra (2006), presentan muchas ventajas en distintos ámbitos lingüísticos, como en la terminografía, la didáctica de lenguas extranjeras, la traducción y la lingüística computacional (Vargas Sierra, 2006). En nuestro trabajo hemos elegido como herramientas didácticas este tipo de corpus, más concretamente, los corpus paralelo chino-español.

Teniendo en cuenta los corpus paralelos que vamos a utilizar en este trabajo (apartado 5), conviene introducir un tipo de corpus muy especial que se ha llegado a denominar “corpuscionario”. Puesto que las fronteras entre el corpus y el diccionario vienen desapareciendo en los últimos tiempos, el neologismo “corpuscionario” (acuñado por Alonso Ramos, 2009) implica la hibridación entre corpus y diccionario. Es decir, se refieren a los diccionarios elaborados a partir de corpus (generalmente de acceso libre y con información textual), o los corpus anotados y complementados por diccionario (enriquecidos con otros tipos de información sintáctica y semántica). Según Alonso Ramos (2009), un ejemplo de estos híbridos podría ser “la reutilización de un corpus

paralelo aunque complementando con otros repertorios lexicográficos” (Alonso Ramos, 2009: 1195). Por ejemplo, como veremos más adelante (véase los apartados 3.2.4.1 y 3.2.4.3), las plataformas famosas como *Linguee*⁹, *Glosbe*¹⁰ y *Reverso*¹¹ muestran dicha interacción entre corpus y diccionario, en otras palabras, son diccionarios bi-/multilingües de libre acceso en línea que están basados en corpus paralelos alineados. El núcleo de este tipo de diccionario son los ejemplos de contextualización en dos o más lenguas, ya que constituye una interfaz de acceso libre a los corpus paralelos. Es decir, nos posibilita acceder a las concordancias paralelas, como hemos visto en el apartado 3.1.2.1. A continuación, para tener una idea más clara y precisa, presentamos el resultado de una búsqueda en el diccionario *Linguee*, que contiene tanto las entradas de diccionario como las traducciones alineadas que incluyen la palabra que buscamos en contexto procedentes de sitios web bilingües y multilingües.

⁹ Un diccionario en línea que recopila las traducciones realizadas por humanos, alineadas con su texto de origen, y tomadas de los sitios web multilingües (Alonso Jiménez, 2013). Cabe mencionar que, a pesar de que en *Linguee*, un “corpuscionario” que tiene mucha popularidad, se incluye el chino, solo se puede combinar con el inglés para formar par de idiomas. Es decir, en nuestro caso, no nos proporciona las concordancias paralelas chino-español. Disponible en: <https://www.linguee.es/>,

¹⁰ Disponible en: <https://es.glosbe.com/>

¹¹ Disponible en: <https://diccionario.reverso.net/>

The screenshot shows the Linguee website interface. At the top, there's a navigation bar with the DeepL logo, 'Traductor', and 'Linguee'. Below this, a language selector shows 'español' and 'inglés'. A search bar contains the word 'ejemplo'. The results section is titled 'Diccionario español-inglés' and lists several translations for 'ejemplo' with example sentences in both languages.

Spanish Word	Spanish Example	English Translation	English Example
ejemplo (sustantivo, masculino) (plural: ejemplos m)			
example s (uso frecuente) (plural: examples)	El ejemplo anterior es perfecto para el libro. La profesora aclaró el concepto con un ejemplo.	The above example is perfect for the book. The teacher clarified the concept with an example.	
instance s (plural: instances)	El último libro del autor es un ejemplo de la nueva tendencia literaria.	The author's latest book is an instance of the new literary trend.	
cue s (plural: cues)	Seguí el ejemplo de mi madre y trabajé duro para alcanzar mi meta.	I took a cue from my mother and worked hard to reach my goal.	
lead s (plural: leads)	El equipo siguió el ejemplo del miembro con más experiencia.	The team followed the lead of the most experienced member.	
model s (plural: models)	Mi madre es un ejemplo de bondad.	My mother is a model of kindness.	

Ilustración 7 Consulta en *Linguee*

En comparación con los diccionarios bi-/multilingües convencionales, el uso de los “corpuscionarios” de tipo *Linguee* muestran unas ventajas importantes en la enseñanza-aprendizaje de lenguas. Por un lado, los diccionarios convencionales han recibido críticas por su función limitada de transcodificación y la poca información textual¹². Mientras que los “corpuscionarios” podrían salvar dicha deficiencia, puesto que nos presenta la correspondencia entre párrafos, oraciones o frases entre la lengua original y la traducción, con la que se podrían satisfacer las necesidades de los estudiantes de la información a nivel sintáctico, pragmático, gramatical, etc. Por otro lado, este tipo de “corpuscionario” tiene asegurada una gran cantidad de datos lingüísticos, ya que suele usar la web como corpus paralelo (véase el apartado 3.2.3). Con respecto a los corpus paralelos chino-español que manejamos en este trabajo, como veremos en el apartado

¹² Por ejemplo, en Hernández (2000) se describe la limitación de los diccionarios bilingües en la enseñanza del ELE.

5, cabe adelantar que *Eshelper*¹³ y *Glosbe*¹⁴ pertenecen a la categoría de los “corpusionarios”, considerados como herramientas con las que los estudiantes se sienten cómodos en su aprendizaje de lenguas (Cruz, 2017).

3.1.4 Aplicaciones didácticas de los corpus en el ámbito del ELE

Puesto que los objetivos de este trabajo son conocer las aplicaciones didácticas de los corpus paralelos (objetivo 1) y desarrollar una propuesta específica para los alumnos sinohablantes (objetivo 3), en este apartado vamos a conocer y reflexionar primero sobre los usos didácticos de los corpus (en general) en la enseñanza de lenguas extranjeras¹⁵. Más adelante (en el apartado 3.2.4) nos centraremos en las aplicaciones de los corpus paralelos en particular.

Con respecto a la investigación sobre la aplicación de los corpus en la enseñanza de lenguas, muchos antecedentes nos han mostrado su uso didáctico desde diferentes perspectivas. Por ejemplo, McEnery y Xiao (2011) desarrollan un estudio exploratorio sobre los tres tipos de usos didácticos de los corpus. Estos autores distinguen el uso indirecto de los corpus, el uso directo de los corpus y el desarrollo de metodologías pedagógicas basadas en corpus (como se verá más adelante, para la elaboración de la propuesta didáctica, es decir, el objetivo 3, nuestro interés reside en el uso directo de los corpus). En primer lugar, en cuanto a los usos indirectos de los corpus para favorecer la didáctica de lenguas, los corpus son muy útiles para elaborar diccionarios y materiales de referencia, pruebas o exámenes, y mejorar la formación de profesores de lenguas. En segundo lugar, en los usos directos de los corpus, una vez adquirido los

¹³ Disponible en: <https://www.esdict.cn/>

¹⁴ Disponible en: <https://es.glosbe.com/>

¹⁵ Más adelante (en el apartado 3.2.4.) nos centraremos en las aplicaciones de los corpus paralelos en particular.

conocimientos necesarios de LC (entendida como una subdisciplina de lingüística igual que la sintaxis o la pragmática), los alumnos tienen acceso a explorar los corpus según sus propias necesidades. De ahí que los corpus favorezcan el aprendizaje de los alumnos en el marco del Data-Driven Learning (Aprendizaje Basado en Datos)¹⁶, es decir, los alumnos tienen oportunidades de explorar y descubrir las reglas ellos mismos. En tercer lugar, los corpus pueden utilizarse para favorecer el desarrollo de metodologías pedagógicas basadas en corpus. Esta última utilidad ha sido ampliamente desarrollada, por un lado, en la enseñanza de lenguas con fines específicos, y, por otro lado, en la investigación sobre la adquisición de lenguas a partir de corpus de aprendices¹⁷. Una vez presentados estos tres usos didácticos, el estudio de McEnery y Xiao (2011) se concentra más en los usos indirectos de los corpus en la enseñanza de lenguas, puesto que consideran que estos usos indirectos son más empleados que los directos.

Dado que uno de los objetivos principales de este trabajo fin de máster es elaborar una propuesta didáctica utilizando los corpus paralelos (objetivo 3), en nuestro caso, chino-español, nuestro enfoque será el de un uso directo de los corpus en el aula. Sin duda, son muchos los usos directos que pueden aportar los corpus en la enseñanza de lenguas. Aquí presentamos algunos, sugeridos por Alonso Pérez-Ávila (2006):

a) Los corpus como fuente de input:

En nuestro trabajo fin de máster, para elaborar una propuesta didáctica (objetivo 3), los corpus nos permitirán disponer de ejemplos reales en el diseño de actividades. Con estos ejemplos se pueden observar las características lingüísticas, por ejemplo, las de

¹⁶ Sobre Data-Driven Learning (DDL), puede verse Tim Johns (2000).

¹⁷ Un corpus de aprendices es un tipo particular de corpus, que recopilan datos escritos u orales producidos por los niños adquiriendo su lengua materna, o por los individuos que están aprendiendo una segunda lengua o lengua extranjera (McEnery y Xiao, 2011). Sobre corpus de aprendices del ELE de L1 chino, véase Lin (2005); Jiménez y Tang (2018); Sun (2019); Calero et al. (2020).

tipo léxico, gramatical, sociocultural, discursivo, etc., de las lenguas. Las muestras reales de una lengua evitan que los ejemplos siempre sean inventados por los profesores que provienen de su idiolecto (Alonso Pérez-Ávila, 2006). Cruz (2017) apoya esta idea y afirma que los corpus empleados ponen a los estudiantes (y también al profesor) ante más muestras de lengua de distintas fuentes. Los estudiantes pueden recibir una cantidad considerable de input real, comprensible y variado. Asimismo, se ofrecen muestras reales en diferentes registros, géneros, dialectos, cronolectos, sociolectos, etc.

b) Los corpus en la gramática inductiva:

Los estudios previos han señalado que el input que recibe el aprendiz permite su elaboración de hipótesis sobre las reglas gramaticales y las estructuras lingüísticas de la L2 (Alonso Pérez-Ávila, 2006). Gracias a los ejemplos de uso abundantes de una palabra en un corpus, los alumnos tienen la oportunidad de observar cómo se utiliza en contexto y elaborar su propia hipótesis, de ahí que los corpus les posibiliten extraer reglas gramaticales ellos mismos (aprendizaje inductivo). En nuestro caso, los corpus paralelos ofrecen los ejemplos no solo en la lengua meta, sino también en la lengua materna de los aprendices. En nuestra opinión, esta particularidad permite que los aprendices puedan entender con más precisión el significado de cada frase, conocer el uso de las palabras con las que no están familiarizados, comparar las reglas gramaticales de las dos lenguas y verificar su hipótesis con más facilidad.

c) Los corpus en el aprendizaje de colocaciones y combinaciones:

Utilizando los corpus, los alumnos disponen de abundantes contextos en los que aparecen las colocaciones y combinaciones de palabras, y pueden llegar a una mayor comprensión de este concepto (Alonso Pérez-Ávila, 2006). Con las consultas de las combinaciones de palabras en el corpus, los alumnos pueden conocer el valor semántico

de las combinaciones léxicas, saber la importancia de aprender el léxico en contexto, y recordar las colocaciones con ayuda de ejemplos reales.

d) Los corpus de textos literarios en clases de literatura:

Para los profesores que se encarguen de las clases de literatura, los corpus formados por textos literarios pueden ser de gran utilidad para realizar estudios estilísticos y temáticos (Alonso Pérez-Ávila, 2006). Cabe mencionar que, en nuestro caso, actualmente hay pocos corpus paralelos chino-español que estén formados por textos literarios. Pero esta utilidad de corpus nos parece de gran potencialidad en la futura enseñanza de la literatura española en China, dado que la traducción al chino de las obras literarias, en nuestra opinión, facilitará la lectura y la comprensión de las obras para los alumnos sinohablantes.

Cruz (2017), por su parte, ofrece una visión alternativa sobre las aplicaciones de corpus en la enseñanza del ELE al aplicar los principales resultados (concordancias, ocurrencias, y listas de frecuencias) de los corpus, los corpus de aprendices, corpus lematizados, y corpus para la enseñanza de lenguas para fines específicos. Entre ellos, vamos a centrarnos en las aplicaciones de la consulta de las concordancias y ocurrencias para apoyar nuestra elaboración de propuesta didáctica (objetivo 3). A partir de las propuestas de Honeyfield (1989) y Cruz (2017), podemos destacar las siguientes aplicaciones de las líneas de concordancias en la clase de ELE:

- a) Trabajar con las combinaciones y las colocaciones, inducir el valor semántico de cada combinación, descubrir las combinaciones posibles.
- b) Aprender la diferencia de uso y significado en diferentes variedades del español y en diferentes tipos o muestras de textos. Por ejemplo, aclarar las dudas sobre la variedad geográfica, resolver dudas futuras parecidas con

autonomía, explorar el tono o estilo en diferentes tipos de textos, comparar la formalidad entre la lengua escrita y oral, etc.

- c) Observar los ejemplos en los corpus paralelos y extraer conclusiones de esta observación de modo inductivo.
- d) Rellenar huecos en las concordancias tras observar los contextos disponibles.
- e) Completar, o adivinar el contexto de las líneas de concordancias.
- f) Hacer ejercicios sobre análisis de discurso a través de las concordancias.
- g) Uso libre de los programas de concordancias para favorecer su expresión, corrección y comprensión escrita.

La aplicación de corpus en la enseñanza de lenguas se está instaurando poco a poco en el aula y, como acabamos de ver, los usos didácticos de los corpus presentan muchas ventajas. Pero, a pesar de todas las posibilidades que hemos mencionado en este apartado, hay que admitir que también existen algunos inconvenientes y problemas a la hora de aplicar los corpus en la enseñanza de lenguas. A continuación, resumimos las ventajas y desventajas principales de la enseñanza-aprendizaje basado en corpus que señala Valverde (2012), complementando con otras lecturas:

Ventajas:

- **Objetividad:** el corpus es un material auténtico basado en lengua real, así que los alumnos pueden tener acceso a muestras reales producidas por los nativos y evitar los ejemplos artificiales y subjetivos.
- **Cantidad ingente de textos:** los corpus nos permiten obtener textos numerosos y variados, con los que podemos trabajar en la enseñanza de lenguas en función de diferentes objetivos y necesidades.
- **Observación del comportamiento de la lengua en su contexto:** La gran cantidad de datos que almacenan los corpus nos brindan contextos culturales

y situacionales de cada término. El docente y los estudiantes pueden acceder a esos datos y observar la lengua en su propio contexto.

- **Acceso en línea:** muchos corpus son de acceso libre y gratuito en línea, lo cual muestra la facilidad de acceso a este tipo de recurso.
- **Autonomía de aprendizaje:** a través del uso directo de los corpus, los alumnos pueden adquirir mucho conocimiento por sí solos, verificar sus hipótesis y extraer sus propias conclusiones de modo inductivo, de ahí que se aumente la autonomía del aprendizaje (Valverde, 2012).

Desventajas:

- **Información excesiva:** un corpus contiene una gran cantidad de información a nivel lingüístico y cultural, la cual puede confundir a algunos aprendientes en su análisis y en su interpretación. Como señala Alonso Pérez-Ávila (2006), los alumnos disponen de gran libertad en el proceso de la utilización de los corpus, por eso será difícil “controlar los contenidos y los objetivos que se quieran conseguir en cada unidad didáctica (Alonso Pérez-Ávila, 2006: 10)”.
- **Complejidad:** desde la perspectiva del docente, los corpus muestran cierta complejidad en el uso tanto por su forma y contenido, como por las herramientas técnicas que tiene que conocer para su utilización en clase. Además, la falta de formación previa del docente y la inexperiencia de los aprendientes se consideran una dificultad y un reto (Valverde, 2012). Y si nos centramos en el contexto de la enseñanza de español en China, la mayoría absoluta de los estudiantes tienen poca experiencia en el uso de un corpus. Por lo tanto, les resultaría difícil manejar este recurso y podrían desorientarse ante grandes datos recopilados en los corpus (Yao, 2018).

Pese a que la aplicación didáctica de los corpus ha recibido algunas críticas por algunas desventajas como las que hemos mencionado arriba, es indudable que la utilización de los corpus en la enseñanza-aprendizaje de lenguas, como complemento, favorece el desarrollo del aprendizaje y es muy beneficiosa para los aprendientes (Valverde, 2012). Asimismo, para que este tipo de utilización sea productiva, beneficiosa y efectiva, se requiere mucho esfuerzo. Por ejemplo, como se verá posteriormente en el apartado 4 y 5, hay que seleccionar bien los corpus y hacer una adaptación adecuada al nivel de los aprendientes. Además, para reducir la complejidad, son necesarias las sesiones introductorias para que los alumnos puedan manejar bien esta herramienta guiados por los profesores. Al mismo tiempo, se puede empezar con los corpus más sencillos y luego trabajar con los corpus en el aula de forma gradual (Valverde, 2012).

3.2 Corpus paralelos y su uso práctico en el ELE

3.2.1 Definición y tipología

Como hemos explicado en el apartado 3.1.3, un corpus puede contener más de una sola lengua. En los corpus bilingües (que incluyen textos de dos lenguas) y multilingües (que contienen más lenguas), los textos del corpus están seleccionados según idénticos criterios en una y otra lengua (Hallebeek, 1999). En tales casos, se encuentran el corpus paralelo y el corpus comparable. Sin embargo, hay que admitir que existe cierta confusión sobre la terminología de los corpus bilingües y multilingües. En este trabajo, optamos por la terminología más extendida y aceptada ampliamente que proponen McEnery y Hardie (2012). Según McEnery y Hardie (2012), generalmente, hay tres tipos de corpus que incluyen más de una lengua, que son los siguientes:

- Tipo A: textos de origen en un idioma acompañados por sus traducciones en dos o más idiomas.
- Tipo B: pares o grupos de corpus monolingües seleccionados según criterios comunes en idiomas diferentes.
- Tipo C: combinación de tipo A y tipo B.

Siguiendo a McEnery y Hardie (2012), denominamos los corpus de tipo A como corpus paralelo y los de tipo B, corpus comparable. Los corpus comparables se refieren a aquellos que incluyen textos monolingües en diferentes lenguas que comparten materia, género, tipo de texto, registro, etc. Los textos de un corpus comparable se seleccionan de acuerdo con criterios comunes en diferentes lenguas. Los corpus paralelos, al contrario, se definen como aquellos que contienen la misma colección de textos en dos o más idiomas, uno de ellos es el idioma original y los otros son la traducción. Como ejemplo, uno de los primeros corpus paralelos es el Hansard Corpus¹⁸, un corpus establecido a principios de la década 1990, y formado por actas del Parlamento canadiense en versión inglesa y francesa.

Siguiendo a McEnery y Xiao (2007), los corpus paralelos pueden clasificarse a partir de dos criterios: los idiomas incluidos y la dirección de traducción. Según los idiomas que incluye, un corpus paralelo puede ser bilingüe (los que constan de textos en dos idiomas) o multilingüe (los que constan de textos en más idiomas). Según la dirección de traducción, se pueden clasificar en unidireccionales, si la dirección de la traducción es unívoca (por ejemplo, textos españoles traducidos al chino), bidireccionales, cuando la traducción se efectúa en los dos sentidos (por ejemplo, textos españoles traducidos al chino y vice-versa), o multidireccionales, cuando un mismo original es traducido a

¹⁸ La versión más actualizada se distribuye a través de <http://www.tsrali.com/Main.aspx?cc=true> abonando una suscripción.

varios idiomas. Utilizamos el siguiente diagrama para presentar la tipología de los corpus bi-/multilingües de manera esquematizada:

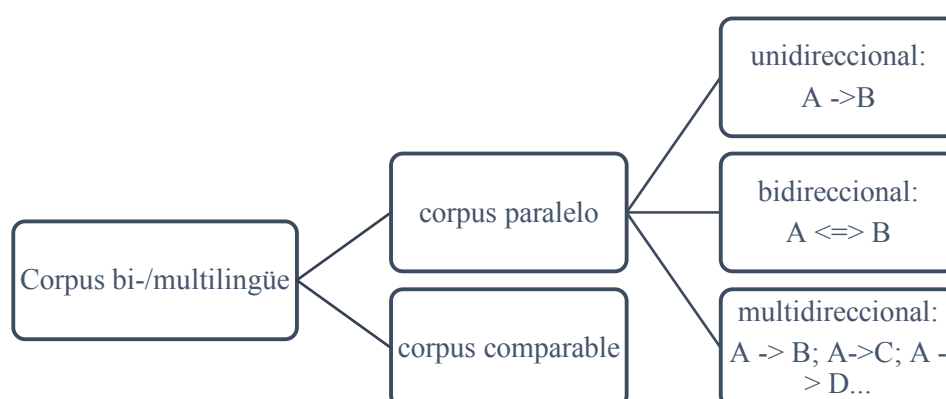


Figura 1 Tipología de los corpus bi-/multilingües, elaboración propia a partir del esquema que propone

Doval (2017)

3.2.2 Diseño y elaboración

Conocer cómo se construye un corpus paralelo nos sirve de gran ayuda saber mejor sus características, su naturaleza y su utilidad (objetivos 1 y 2). En este apartado, se describirá el diseño y la elaboración de un corpus paralelo siguiendo principalmente a Doval (2017), y complementando su propuesta con otras lecturas. Esta autora nos proporciona información detallada sobre los pasos y numerosos aspectos a considerar en la construcción del corpus paralelo bilingüe PaGeS¹⁹, que se puede usar para múltiples fines, especialmente para la investigación en lingüística contrastiva, la traducción y la enseñanza de lenguas extranjeras (Doval, 2017). El motivo de basarnos en este estudio consiste en que el corpus PaGeS es el proyecto original del corpus PaChEs (PAralelo-CHino-Español), un corpus paralelo chino-español en construcción actualmente. Este proyecto, en el que realicé mis prácticas del máster ELEAP, está a

¹⁹ Un corpus paralelo español-alemán, disponible en: <https://www.corpuspages.eu/corpus/about/about?lang=en>

cargo de un grupo de investigación formado por los profesores y estudiantes de la Universidad de Santiago de Compostela, la Universidad de Barcelona y la Fundación de Instituto Confucio de Barcelona. Los corpus PaGes y PaChEs, dirigidos por el mismo equipo de investigación, forman parte de un proyecto de corpus paralelos de español con diversas lenguas iniciado por la Universidad de Santiago de Compostela. Por lo tanto, a nuestro juicio, este estudio nos ofrecerá pistas importantes sobre la constitución de un corpus paralelo chino-español. Según Doval (2017), en el proceso de elaboración de un corpus paralelo, se destacan las tres partes principales: la compilación de textos y su preprocesado, la alineación de los textos y la anotación documental y lingüística.

a) Compilación de textos y su preprocesado

Un corpus debe mostrar el uso real de una lengua y debe tener cierta representatividad, como hemos explicado en el apartado 3.1.1. Por ello, la selección de los textos debe seguir algunos criterios específicos a partir de distintos propósitos para el que se crea (Doval, 2017). Por su parte, Castillo Rodríguez (2009) propone los siguientes criterios de diseño en la fase de recopilación de datos: el objetivo de la construcción; el tamaño; el medio; la temática; el tipo textual; la autoría; la fecha de publicación; y, por último, las lenguas. En el caso de la elaboración de un corpus paralelo para la enseñanza de las lenguas y las investigaciones lingüísticas contrastivas, de acuerdo con Doval (2017), hay que tener en cuenta los siguientes aspectos en la fase de la compilación de textos: En primer lugar, es importante la cantidad y la calidad de los textos. Es decir, por una parte, hay que asegurar la suficiente obtención del material directo bilingüe de las dos lenguas, por otra parte, es importante la calidad del material, es decir, la calidad de los textos originales y de su traducción. Por ejemplo, es recomendable acudir a los materiales publicados por editoriales reconocidas, donde los originales y las

traducciones han pasado un estricto control de calidad. Así que, para evaluar los corpus paralelos chino-español disponibles en línea en el apartado 4 y 5, vamos a considerar la cantidad y la calidad como un criterio fundamental. En segundo lugar, cabe subrayar la importancia del equilibrio de los datos en cuanto a la dirección de traducción, ya que tanto para las investigaciones de lingüística contrastiva como para las de traducción se estudia el proceso mismo de traducción. En tercer lugar, en el caso de la recopilación de obras literarias recientes, hay que tener en cuenta que los textos siempre están protegidos por derechos de autor, lo cual consiste en uno de los factores que limitan la disponibilidad de los corpus paralelos. Por esta razón, cada vez más investigadores se inclinan a utilizar los datos libres de derechos de autor y excluir los usos comerciales de los corpus. En nuestro caso, como se verá más adelante, en el apartado de metodología (apartado 4 y 5) vamos a analizar si los corpus que manejamos en este trabajo cumplen estos requisitos.

Una vez seleccionados y digitalizados, los textos se someten a un proceso manual para prepararlos para lograr el paralelismo y obtener el mejor resultado de la alineación automática. Este preprocesado de textos consiste en básicamente la eliminación de textos no correspondientes, notas del autor o traductor, imágenes y todos los fragmentos que no tienen la correspondencia en la otra versión. Posteriormente se revisan y corrigen ambas versiones, puesto que a veces se producen unos errores en el proceso de digitalización, por ejemplo, espacios redundantes, erratas tipográficas, usos inadecuados ortográficos, etc.

b) Alineación de textos

Un paso crucial en la construcción y elaboración de un corpus paralelo es la alineación (Doval, 2017). En palabras de Castillo Rodríguez (2009: 7), la alineación es “la

reestructuración de los textos de tal forma que se pueda establecer una correspondencia entre los párrafos, oraciones y/o palabras de los textos involucrados”. El conjunto de dichas correspondencias del original y la traducción forman el bitexto alineado. En cuanto a los niveles de la segmentación, las unidades de alineación pueden ser párrafos, frases o palabras. Entre ellos, lo más habitual actualmente es la alineación a nivel oracional, tal como ilustra el siguiente ejemplo extraído del corpus PaChEs²⁰:

ID	ES	ZH
6101-0000{1}	SIERVOS DE LA TIERRA	{1} 土地的奴隶
6101-0000<1>		<1>
6101-0000¿	Año 1320	1320年
6101-0000^	Masia de Bernat Estanyol	柏纳·艾斯坦优农庄
6101-0000\$	Navarres, Principado de Cataluña	纳瓦克雷斯, 加泰罗尼亚王国
6101-0000€	En un momento en el que nadie parecía prestarle atención, Bernat levantó la vista hacia el nítido cielo azul.	趁着大家不注意, 柏纳抬头望了望蔚蓝晴空。
6101-0000€	El sol tenue de finales de septiembre acariciaba los rostros de sus invitados.	九月底的和煦暖阳轻抚着宾客们的脸庞。
6101-0000;	Había invertido tantas horas y esfuerzos en la preparación de la fiesta que sólo un tiempo inclemente podría haberla deslucido.	他投注了许多时间和心力, 大费周章地准备了这样一场盛宴, 事事完备, 只怕天公不作美。
6101-0000€	Bernat sonrió al cielo otoñal y, cuando bajó la vista, su sonrisa se acentuó al escuchar el alborozo que reinaba en la explanada de piedra que se abría frente a la puerta de los corrales, en la planta baja de la masía.	柏纳面带微笑望着初秋的蓝天, 过了半晌, 当他看见农庄前的广场上挤满了兴高采烈的宾客时, 脸上的笑容更灿烂了。

Ilustración 8 Muestra de corpus *PaChEs*

La alineación consta de dos tareas, una previa de segmentación de los textos y la posterior, la vinculación de esos segmentos a partir de la correspondencia entre las dos lenguas para formar bisegmentos, es decir, pares alineados de segmentos (Doval, 2017). Cabe mencionar que la segmentación se realiza tanto en la lengua original como en la traducción independientemente, por consiguiente, se vinculan los segmentos resultantes correspondientes.

Con respecto a los cuantiosos algoritmos existentes para la alineación oracional, se pueden clasificar en tres grupos a partir de su metodología: enfoques basados en la longitud de los segmentos, enfoques de correspondencia léxica y enfoques híbridos (Doval, 2017). Es decir, en el primer enfoque se alinea una oración cuando la lengua

²⁰ Un corpus paralelo chino-español en construcción actualmente. El proyecto está a cargo de un grupo de investigación formado por los profesores y estudiantes de la Universidad de Santiago de Compostela, la Universidad de Barcelona y la Fundación de Instituto Confucio de Barcelona.

fuente contiene un cierto número de caracteres o palabras y en la lengua meta el número de estos es similar. El segundo, en cambio, utiliza un método basado en léxico y usa los diccionarios bilingües o similitudes formales entre las palabras en el proceso de la alineación. El último, los métodos híbridos combinan los dos primeros enfoques. Por su parte, Braune y Fraser (2010: 81) argumentan que los métodos híbridos “constantly provides high quality alignments”.

En cuanto a la herramienta de auto-alineación, Doval (2017) confirma que, comparada con otras herramientas informáticas²¹, la aplicación alineadora LF-Aligner²² alcanzó la mayor precisión en varias pruebas. Utiliza tanto la longitud de la oración como las correspondencias léxicas para tener la propuesta de alineación final. Ahora bien, durante el proceso de traducción, las oraciones pueden ser divididas, insertadas, reordenadas o resumidas por el traductor para que el texto en la lengua meta sea natural. Y todo esto provoca desafíos importantes para la auto-alineación. Después del proceso automático, es necesario revisar y validar manualmente esta alineación para obtener resultados fiables y precisos, normalmente con una tasa de error inferior al 0,5% (Doval, 2017).

Sin embargo, comparados a los corpus paralelos que incluyen lenguas europeas, los que incluyen las lenguas no alfabéticas como el chino son muy poco estudiados, debido a la diferencia fundamental entre los sistemas lingüísticos que causa cierta dificultad y problema conceptual y computacional en el proceso de alineación (Wang, 2001). Cabe mencionar que, en nuestro caso, el chino y el español son dos idiomas extremadamente diferentes desde el punto de vista léxico, sintáctico, discursivo, etc. Por ejemplo, puesto

²¹ Alineadores como CWB-align (<<http://cwb.sourceforge.net/>>); Vanilla (<<http://nl.ijs.si/telri/Vanilla/>>); Abbyy Aligner (<<https://www.abbyy.com/en-eu/aligner/>>).

²² Se puede obtener esta aplicación a través de <https://sourceforge.net/projects/aligner/>

que el chino es un idioma más compacto que el español, es frecuente que el número de los caracteres en chino sea menor que el número de las palabras en español a la hora de expresar la misma información. Así que, como ejemplo, durante el proceso de la alineación del corpus paralelo chino-español PaChEs, la alineación mecánica previa a través de la aplicación informática (LF-Aligner) se guía por puntos de cierre [.), tales como los de frase, o bien los signos de interrogación[?) y exclamación[!) de cierre para formar bi-textos. No obstante, tras este primer alineamiento, se observan frecuentemente diversas incidencias causadas por las formas distintas de distribución de la información y las divergencias de los usos sintácticos del chino y del español. De ahí que se exija la revisión y el reagrupamiento manualmente de los fragmentos una vez obtenida la primera propuesta de texto alineado en forma de bisegmentos²³.

c) Anotación

Siguiendo a McEnery y Hardie (2012), un corpus contiene básicamente los siguientes tres tipos de información añadida para la investigación de los datos en el corpus: los metadatos, el marcado textual y la anotación lingüística. Más concretamente, los metadatos se refieren a la información externa relacionada con la obra original, la traducción y otros datos bibliográficos. Este tipo de información nos permite, posteriormente, recuperar información relevante del corpus. El marcado textual, en cambio, proporciona la división interna de los textos tales como capítulos o partes a fin de localizar una cadena de texto dentro de una obra. Por último, la anotación lingüística indica un cierto análisis lingüístico a diferentes niveles de los datos. Según Doval (2017), en la construcción del corpus PaGeS, se utiliza el nombre de archivo para proporcionar una lista de metadatos que incluye información sobre el autor y/o

²³ Información obtenida del *Protocolo para constituir el corpus chino-español del Proyecto PaChEs* (elaborado por la Universidad de Santiago de Compostela, la Universidad de Barcelona y la Fundación de Instituto Confucio de Barcelona).

traductor, título, fecha, información sobre la publicación, lengua original y la lengua de la versión, etc. Luego, se introducen manualmente los marcados textuales en los bitextos. Después, se separan de nuevo los textos por lengua para proceder su anotación respectivamente (Doval, 2017). Como se ha visto en el apartado 3.1.2.3., la lematización es fundamental a la hora de trabajar con corpus, por eso, para la lematización del corpus PaGeS, se recurre a unos programas informáticos para la agrupación de las formas flexivas de las palabras por lemas y su etiquetación en categorías gramaticales.

3.2.3 La web como corpus paralelo

En la mayoría de las ocasiones, los corpus paralelos tienen un tamaño respectivamente limitado y es habitual que las concordancias paralelas mostradas sean insuficientes para extraer conclusiones fiables. Ante este problema, una solución es utilizar la web como corpus. Como se verá más adelante en el apartado 4, una gran parte de los corpus paralelos chino-español disponibles en línea son compilados a través de Internet, utilizando los textos y datos extraídos de los sitios web bi-/multilingües. Así en un corpus tan enorme, habrá información suficiente para “aprender las propiedades de un gran número de palabras” (Gelbukh, Sidorov, y Chanona-Hernández, 2002). A diferencia del proceso de la constitución de un corpus paralelo como hemos presentado en el apartado 3.2.2, la recopilación de los corpus paralelos a partir de la web requiere varios instrumentos técnicos y sistemas informáticos. Estos pueden incluir las herramientas para localizar, extraer, y filtrar los datos paralelos en los sitios web (*p. ej.* STRAND [Structural Translation Recognition, Acquiring Natural Data]²⁴, WPDE²⁵

²⁴ Sobre STRAND, véase Resnik y Smith (2006).

²⁵ Sobre WPDE, véase Zhang, Wu *et al.* (2006)

[Web Parallel Data Extraction], PCMS [Parallel Corpus Mining System]²⁶, etc.), los sistemas de alineación, los etiquetadores, etc.

Cabe adelantar que, entre los corpus seleccionados en nuestro trabajo, cuatro de ellos se compilan a partir de los datos de la web²⁷. Los corpus de este tipo son fáciles de manejar y contienen gran cantidad de datos bi-/multilingües, pero también presentan algunos problemas a la hora de utilizarlos en el aula. Uno de estos consiste en “la baja calidad de los contextos encontrados” (Gelbukh, Sidorov, y Chanona-Hernández, 2002), como veremos en el apartado 5.6. Por ejemplo, los textos pueden ser producidos por autores poco cultos, la alineación automática sin revisión puede causar incidencias, los textos pueden ser mal segmentados, etc.

3.2.4 Aplicaciones didácticas de los corpus paralelos

Si nos centramos en las aplicaciones de los corpus paralelos, Doval (2017) propone que generalmente se distinguen los siguientes cinco grandes campos, con sus usuarios específicos respectivamente: (a) la investigación básica en lingüística contrastiva y translatología, (b) la lexicografía, (c) la traducción, (d) la enseñanza de lenguas extranjeras y de la traducción y (e) el procesamiento del lenguaje natural, especialmente la traducción automática (Doval, 2017). En realidad, los corpus paralelos resultan muy útiles en la investigación sobre la traducción tanto para la formación de traductores como para la mejora de los resultados de sistemas de traducción automática. En el primer caso, este tipo de corpus es una fuente de datos directa para las máquinas, y en el otro, los corpus paralelos contienen evidencia de cómo un traductor resuelve problemas que se producen en la práctica de la traducción (Hallebeek, 1999). En este

²⁶ Sobre PCMS, véase Li y Liu (2008).

²⁷ Los cuatro corpus son: TAUS Data Cloud, Glosbe, Eshelper y ReversoContext.

sentido, la aplicación de los corpus paralelos en los estudios de traducción y el entrenamiento de los programas automáticos de ayuda para la traducción se investiga desde hace años. Pero los corpus paralelos también se están empleando cada vez más en el ámbito de la enseñanza de lenguas (Cruz, 2017), un terreno en el que este tipo de corpus presenta ventajas como las siguientes:

➤ **Mostrar la correspondencia entre diferentes lenguas**

Como sostiene Barlow (2000), utilizar los textos paralelos permite a los aprendices de lenguas investigar directamente la correspondencia entre las palabras o estructuras en dos lenguas. Además, en palabras de Simões (2004: 7), a través de la lectura de los textos paralelos, los alumnos pueden “try to understand the translated sentence and mentally align concepts and structures with the original one (Simões, 2004)”, lo cual favorece su aprendizaje de lenguas. En el apartado 6.2, veremos más concretamente cómo la correspondencia facilitará el aprendizaje de español para los alumnos sinohablantes.

➤ **Ofrecer la evidencia de contrastividad:**

De acuerdo con Kenning (2010), aparte de las ventajas que aportan los corpus en general, los corpus paralelos permiten a los estudiantes acceder la evidencia contrastiva que destaca las similitudes y divergencias entre las diferentes lenguas y les sirven de ayuda para procesar los datos lingüísticos de la lengua meta. Además, siguiendo a Frankenberg-Garcia (2004), el uso directo de las concordancias paralelas en el aula ayudará a los alumnos a ser conscientes de sus errores y problemas interlingüísticos en su producción de la lengua meta causados por las interferencias de su lengua materna.

➤ **Autonomía de aprendizaje**

Además, el acceso a la traducción, tanto la completa como la parcial, posibilita que los estudiantes estén expuestos a las ocurrencias que contienen vocabulario o frases que no conocen. De ahí que se favorezca el auto-aprendizaje de los estudiantes y se facilite el Data-Driven Learning (Aprendizaje Basado en Datos)²⁸ (Kenning, 2010).

Con el objetivo de conocer y reflexionar sobre las aplicaciones didácticas de los corpus paralelos en la enseñanza de lenguas (objetivo 1), a continuación, se verá que los estudios distinguen cuatro tipos de usos de los corpus paralelos en este ámbito: uso directo de las concordancias paralelas en el aula, uso indirecto de los corpus paralelos para la elaboración de materiales didácticos (sobre el uso directo e indirecto de los corpus en general en la enseñanza de lenguas, véase el apartado 3.1.4.), uso como herramienta de consulta para mejorar la expresión escrita y uso como una base de la investigación en lingüística contrastiva.

3.2.4.1 Uso directo en el aula de la consulta de concordancias paralelas

El uso directo de los corpus paralelos en el aula ha sido investigado por muchos autores, especialmente en lo concerniente a los corpus que incluyen el inglés y desde la década de 1990. Un ejemplo específico del uso directo es el estudio de Barlow (2000). Utilizando los datos bilingües inglés-francés, Barlow (2000) propone que se puede sensibilizar a los estudiantes sobre la manera de expresar el significado equivalente en diferentes lenguas, gracias a los abundantes contextos que proporcionan. Por ejemplo, a través de la consulta de los pronombres reflexivos de dos lenguas, los alumnos pueden observar que las formas reflexivas en inglés no se traducen necesariamente a una forma reflexiva en francés, e incluso para algunas formas no existen equivalentes similares en

²⁸ Sobre Data-Driven Learning (DDL), puede verse Tim Johns (2000).

otra lengua. Otra propuesta de Barlow (2000) en este mismo estudio consiste en la observación de una lista de concordancias de una palabra determinada para ver que, en otra lengua, se corresponde a varias unidades léxicas diferentes en función de los distintos contextos. Por ejemplo, con los corpus paralelos inglés-español se propone que, gracias a la observación de los abundantes contextos de la palabra *for* en inglés (que puede corresponder a *por* o *para* en español), los alumnos pueden elaborar hipótesis sobre la traducción de esta preposición. El mismo autor señala que el uso de los corpus y concordancias permite que todos, tanto los lingüistas teóricos como los aprendices de L2, se conviertan en investigadores (Barlow, 2000).

Otro ejemplo parecido es un estudio basado en un corpus inglés-alemán que propone St. John (2001). En este estudio se describe la utilidad de los corpus paralelos con alumnos de nivel inicial. Las tareas propuestas en este caso consisten en la comparación entre la lengua meta y la lengua primera para extraer conclusiones de esta observación. Por ejemplo, a partir de la consulta de la palabra *also* en alemán, que puede ser una partícula o un adverbio dependiendo de la sintaxis y el contexto, y que es un *falso amigo* entre estas dos lenguas, los alumnos pueden observar la función, la posición y el uso como modificador de esta palabra e intentar derivar independientemente las reglas que rigen su uso.

Aparte de los dos casos mencionados arriba, Zanettin (1994) propone tres tipos de actividades que se pueden llevar a cabo con los corpus paralelos italiano-inglés, que son las actividades guiadas por docentes, las semi-guiadas y las actividades sin guía del docente. Por su parte, Frankenberg-Garcia (2005) nos presenta los usos diferentes de las concordancias paralelas y monolingües en las distintas situaciones en el aprendizaje de L2. Afirma que el uso de las concordancias paralelas portugués-inglés en el aula

facilitará la observación de diferentes equivalencias en distintos contextos, la comprensión de los conceptos culturales, la reflexión sobre los errores interlingüísticos, etc. Asimismo, Wang (2001) realiza un estudio sobre el aprendizaje del léxico utilizando las concordancias paralelas chino-inglés, lo cual nos parece especialmente interesante dado nuestro objetivo de investigación porque se trata de un estudio sobre las aplicaciones pedagógicas de un corpus paralelo que incluye el chino. Wang (2001) propone que los alumnos sinohablantes de nivel intermedio de inglés pueden extraer conclusiones sobre cómo usar el adverbio *now* en inglés en diferentes contextos. Tras observar las concordancias paralelas y comparar las estructuras diferentes en dos lenguas, los alumnos son capaces de descubrir por sí mismos los matices de significado en función de los distintos contextos y evitar producir las estructuras inapropiadas en la lengua meta.

Frente a estos trabajos sobre las aplicaciones didácticas de los corpus paralelos que incluyen la lengua inglesa, hay que admitir que hasta ahora los estudios sobre las aplicaciones de corpus paralelo en la enseñanza del ELE han sido más bien modestos. Cabe mencionar aquí el estudio de Doval (2018), que ilustra muy bien la posibilidad del uso del corpus PaGeS, un corpus paralelo español-alemán en el aula de ELE. Dicho corpus pretende ser un recurso útil para los estudiantes avanzados de español o de alemán, para que dispongan de diversos contextos de uso y ejemplos de traducción (Doval, 2018). Además, Olímpio de Oliveira y Penadés (2016) nos presentan cómo utilizar los “corpuscionarios” de tipo *Linguee* (sobre el “corpuscionario”, véase el apartado 3.1.3), un tipo especial de corpus paralelos, en la enseñanza-aprendizaje de las unidades fraseológicas en el aula de ELE: las actividades didácticas que presentan en su estudio consisten en la elaboración de un diccionario bilingüe de las locuciones, un proyecto en el que trabajan los alumnos durante todo el curso bajo las instrucciones por

parte del docente a partir de la aplicación *Linguee*. Como se ha visto en el apartado 3.1.3, las aplicaciones como *Linguee* son herramientas consideradas corpus paralelos gratuitos de fácil acceso (Alonso Jiménez, 2013). Es decir, son diccionarios que “explican” el significado de una palabra, o una combinación mostrándola en su contexto en varias lenguas y, generalmente, se usa la web como corpus multilingüe. Se observa que los “corpuscionarios” de tipo *Linguee* se utilizan habitualmente para el desarrollo de la competencia léxico-semántica (p. ej. Alonso Jiménez, 2013; Olímpio de Oliveira, 2015; Olímpio de Oliveira y Penadés, 2016; Egido y Meliss, 2017 y Martínez, 2018) y, en menor medida, para el desarrollo de la competencia discursiva (Flores Acuña, 2001) de los aprendices del ELE.

Sobre las aplicaciones de los corpus paralelos chino-español en la enseñanza del ELE, cabe destacar los estudios de Lu y Cheng (2016) y Cao, da Cunha e Iruskieta (2018). Lu y Cheng (2016) sugieren que la utilización del corpus CPEIC²⁹ para los estudiantes sinohablantes podrá “facilitar la manipulación de la lengua meta (español), a través de las experiencias y conocimientos de su primera lengua extranjera (inglés) y su lengua materna (chino)” (Lu y Cheng, 2016, 133). Además, propone que los aprendices pueden observar los ejemplos de la estructura “ser/estar+adjetivo” y las traducciones entre las tres lenguas para aprender el uso de estos dos verbos copulativos y las diferencias detalladas en distintos contextos (Lu y Cheng, 2016). Por su parte, Cao *et al.* (2018) crean un corpus paralelo chino-español³⁰ con el fin de favorecer los análisis comparativos discursivos, la investigación sobre la traducción y el aprendizaje del chino y el español. En este estudio, los autores plantean que, comparando y observando la información morfosintáctica y discursiva, los estudiantes sinohablantes podrán ser conscientes de los géneros gramaticales (masculino y femenino) y aprender múltiples

²⁹ Disponible en <http://140.116.245.146/cpeic/TriLin.html?role=teacher>

³⁰ RST Spanish-Chinese Treebank, disponible en <http://ixa2.si.ehu.es/rst/zh/>

marcadores discursivos en el español. Para poder manejar bien este corpus, se requiere un nivel de español mínimo de B2 (nivel intermedio).

Se puede observar que, en el uso directo de los corpus paralelos, es habitual centrarse en las concordancias en lugar del uso estadístico de los corpus. Según Frankenberg-García (2004), es recomendable utilizar los corpus paralelos en el aula cuando el alumnado comparte la misma L1 y el profesor conocer suficientemente bien esta L1. Además, cabe destacar que, en la mayoría de los estudios concretos que hemos mencionado arriba, se confirma la importancia del papel del docente, puesto que los docentes deben familiarizarse con los corpus, recibir la formación de LC previamente y mantener el interés de los alumnos. Además, la utilización de esta herramienta debe adaptarse en función del objetivo, el nivel y la necesidad de los estudiantes. En resumen, todos los estudios mencionados en este apartado nos han ofrecido diferentes enfoques y métodos a la hora de usar los corpus paralelos en el aula, los cuales nos parecen inspiradores y enriquecedores para alcanzar nuestro objetivo de elaborar una propuesta didáctica centrándonos en el uso directo de los corpus paralelos chino-español (objetivo 3).

3.2.4.2 Uso indirecto de los corpus paralelos para la elaboración de materiales didácticos

Aunque en la propuesta de nuestro trabajo no nos vamos a centrar en el uso indirecto de los corpus paralelos, conviene describirlo para tener una visión más global sobre las aplicaciones didácticas de los corpus. Cuando se habla del uso indirecto, los corpus se consideran una fuente ideal para la elaboración de materiales didácticos y para el diseño de ejercicios auténticos, ya que reflejan el uso real de la lengua (Doval, 2017). En este

sentido, los corpus paralelos se consideran como una herramienta complementaria productiva a la hora de crear libros de texto y diseñar cuadernos de ejercicios (Doval, 2017). Cabe mencionar que este tipo de uso se considera muy ventajoso en el ámbito de la enseñanza de lenguas con fines específicos. Como indican Lüdeling y Kytö (2008), un corpus paralelo, sobre todo si contiene textos de especialidad, es adecuado para servir de base para la elaboración de materiales didácticos y manuales de referencia como diccionarios bilingües, puesto que en este tipo de corpus se muestra la equivalencia y correspondencia. A modo de ejemplo, Chang (2004) empezó a crear un corpus³¹ que nos interesa especialmente, que incluía el par de lenguas chino-inglés desde 2001, con la intención de elaborar un diccionario bilingüe chino-inglés.

3.2.4.3 Uso de los corpus paralelos como herramienta de consulta para mejorar la expresión escrita

Frankenberg-Garcia (2004) señala que la consulta tipo *self-access* de los corpus paralelos puede ser de gran ayuda para los alumnos durante las tareas de escritura en la lengua meta, puesto que, por un lado, los corpus paralelos ofrecen tanto el significado de las palabras desconocidas para los alumnos como los contextos donde se usan las palabras adecuadamente, y, por otro lado, les permite observar en ocasiones determinadas unidades léxicas que no tienen equivalente directo en otra lengua. Debido a que la consulta en este caso se inicia por los alumnos mismos según su propia necesidad, este tipo de utilización favorecerá significativamente su aprendizaje de la lengua meta. Bluemel (2014), con el fin de comprobar el valor positivo de incorporar los corpus paralelos en la enseñanza de lenguas extranjeras, llevó a cabo un

³¹ Corpus CCL.

experimento en el que se introduce un corpus paralelo chino-inglés³² como una herramienta de consulta para que los alumnos de chino de nivel inicial puedan mejorar su comprensión lectora y expresión escrita. El resultado puso de manifiesto que los alumnos se sienten cómodos con este tipo de recurso y están motivados por poder utilizar esta herramienta en su aprendizaje. Sin duda alguna, la consulta de tipo *self-access* a los “corpuscionarios” de tipo *Linguee* (sobre el “corpuscionario”, véase el apartado 3.1.3) también favorecerá la expresión escrita de los estudiantes (Buyse, 2014) y les ayudará a aprender los términos en lenguajes de especialidad (Buyse y Verlinde, 2013; Cruz, 2015), puesto que, a través de la observación de las concordancias paralelas (sobre las concordancias paralelas, véase el apartado 3.1.2.1) de una palabra de manera alineada, los usuarios pueden llegar a entender su significado y uso de manera inductiva (Cruz, 2017).

3.2.4.4 Uso de los corpus paralelos como base para la investigación en lingüística contrastiva

Debido a que en los corpus paralelos se muestran los mismos textos en diferentes lenguas de forma alineada, se considera una herramienta muy productiva para desarrollar análisis contrastivos de distintos fenómenos lingüísticos en diferentes lenguas. En la entrada “análisis contrastivo” del *Diccionario de términos clave de ELE*³³, se confirma que, una vez analizadas las similitudes y diferencias entre la lengua meta y la lengua materna, se puede producir una repercusión en la didáctica tanto en las prácticas en el aula como, muy especialmente, en la producción de materiales de enseñanza, puesto que se partía de la base de que los errores producidos por los alumnos eran el resultado de las interferencias de la lengua materna en el aprendizaje de la nueva

³² Parallel corpus teaching tool, disponible en: <http://www.parallelcorpus.com>

³³ Disponible en: https://cvc.cervantes.es/ensenanza/biblioteca_ele/diccio_ele/indice.htm

lengua. Por lo que respecta al tema de este TFM, en los últimos años, el estudio contrastivo entre chino y español a partir de corpus ha suscitado el interés de algunos profesores y estudiantes de español en China. Por ejemplo, tal como exponen Lu y Cheng (2016), Lu y Lu (2016) y Lu *et al.* (2017), se ha utilizado un corpus paralelo español, inglés y chino³⁴ para facilitar varios estudios contrastivos de las tres lenguas sobre los pronombres de complemento directo e indirecto, los verbos copulativos, los tiempos y las voces gramaticales, etc. Por su parte, Luo (2017) publicó su análisis contrastivo sobre las características sintácticas y léxicas de chino, inglés y español en el discurso político, basándose en un corpus paralelo multilingüe³⁵. Ahora bien, como dice Kenning (2010: 492): “the original text and the translation process are bound to exert some influence on the target text, thereby introducing a degree of distortion that affects the reliability of the findings”. Es decir, es inevitable que el texto original y el proceso de traducción ejerzan cierta influencia en el texto traducido, introduciendo así alguna distorsión que afecta la fiabilidad de los resultados. Por tanto, McEnery y Xiao (2007) señalan que, para los estudios contrastivos, es recomendable utilizar los corpus paralelos bidireccionales en combinación con los corpus comparables³⁶ para obtener conclusiones fiables.

En definitiva, una vez presentadas las aplicaciones didácticas de los corpus paralelos, se observa que, hasta la fecha, ha sido relativamente poco estudiada la aplicación de los corpus paralelos dirigida a los aprendices sinohablantes. Para contribuir a llenar este vacío, en primer lugar, en el apartado 4 vamos a enumerar los corpus paralelos chino-español existentes en línea, justificar nuestra selección de unos corpus sobre los que

³⁴ Corpus Paralelo de Español, Inglés y Chino (CPEIC). Disponible en <http://140.116.245.146/cpeic/TriLin.html?role=teacher>

³⁵ Un corpus paralelo del tamaño pequeño creado especialmente para esta investigación, lo cual recopila la versión china, inglesa y español del informe sobre la Labor del Gobierno en 2016.

³⁶ Como hemos dicho en el apartado 3.2.1, los corpus comparables recopilan textos que se seleccionan de acuerdo con criterios comunes en diferentes lenguas.

trabajaremos en este TFM y presentar unos criterios de análisis de los corpus. Y a continuación, en el apartado 5 haremos una descripción de cada uno, para poder elaborar en el apartado 6 unas propuestas didácticas para los alumnos de español cuya lengua materna es el chino.

4. METODOLOGÍA

Como el objetivo principal de este trabajo fin de máster es elaborar una propuesta didáctica utilizando los corpus paralelos para los alumnos sinohablantes (objetivo 3), debemos seleccionar los corpus adecuados para presentar en la clase, y analizar cómo y cuándo utilizar los diferentes corpus siguiendo una serie de criterios (objetivo 2). En nuestro trabajo, vamos a utilizar estos cinco corpus: *TAUS Data Cloud*, *Glosbe*, *Eshelper*, *ReversoContext* y *CPEIC*. En primer lugar, vamos a justificar nuestra selección y presentar los criterios de análisis de los corpus.

4.1 Justificación de selección de los corpus

Ante todo, cabe aclarar que en nuestro trabajo nos vamos a centrar en los corpus paralelos nativos y no nos ocuparemos de los corpus paralelos de aprendices³⁷. Tras utilizar los motores de búsqueda³⁸ y consultar las recopilaciones de los profesionales³⁹ y los estudios relacionados, observamos que la mayoría de los corpus paralelos nativos chino-español existentes hasta la fecha se han compilado en el marco de investigaciones de lingüística computacional⁴⁰ y generalmente no son fáciles de manejar. En cambio, los corpus centrados en las aplicaciones didácticas para la enseñanza de lenguas son relativamente pocos. Con respecto a los corpus paralelos nativos chino-español de fácil

³⁷ Como se ha visto en el apartado 3.1.4, un corpus de aprendices es un tipo particular de corpus, que recopilan datos escritos u orales producidos por los niños adquiriendo su lengua materna, o por los individuos que están aprendiendo una segunda lengua o lengua extranjera (McEnery y Xiao, 2011).

³⁸ Por ejemplo, Google, Baidu, Bing, etc.

³⁹ Por ejemplo, la página personal de Chelo Vargas Sierra, disponible en: <https://personal.ua.es/es/chelo-vargas/enlaces/corpus-paralelos-en-linea.html>, la página de Tagpacker de Mar Cruz Piñol, disponible en <https://tagpacker.com/user/mar.cruz.pinol?t=corpus.paralelos%255Fy%255Fmultilingues>, etc.

⁴⁰ Sobre los corpus paralelos que incluyen el chino y el español centrados en la lingüística computacional como el procesamiento de lenguaje natural y la traducción automática, véase Banchs *et al.* (2006); Wang *et al.* (2008); Costa-Jussà *et al.* (2012); Ling *et al.* (2013); Centelles y Costa-Jussà (2014); Liu *et al.* (2018).

acceso en línea que hemos encontrado, se pueden clasificar básicamente en los siguientes tres grupos:

- Grupo A: los repositorios de traducciones humanas que presentan el texto original y la traducción de forma alineada, y los “corpuscionarios” (véase el apartado 3.1.3) en línea de tipo *Linguee*. Generalmente los textos proceden de sitios web bilingües o multilingües (véase el apartado 3.2.3), y han pasado por un procedimiento altamente automatizado. En este grupo se encuentran los siguientes:
 - *MyMemory* (<https://mymemory.translated.net/>)
 - *TAUS Data Cloud* (<https://data-app.taus.net/>)
 - *Glosbe* (<https://es.glosbe.com/>)
 - *Eshelper* (<https://www.esdict.cn/>)
 - *ReversoContext* (<https://context.reverso.net/traduccion/>)
- Grupo B: los corpus paralelos que suelen servir como la base y la procedencia de los textos de los corpus de grupo A, como los siguientes:
 - *The Holy Bible* (Resnik, Olsen y Diab, 1999) (<http://users.umiacs.umd.edu/~resnik/parallel/bible.html>)
 - *Corpus Paralelo de las Naciones Unidas* (Ziemski, Junczys-Dowmunt y Pouliquen, 2016) (<https://conferences.unite.un.org/uncorpus/>)
 - Corpus en *OPUS* (Tiedemann, 2016) (<http://opus.nlpl.eu/>)
- Grupo C: los corpus paralelos establecidos por un grupo de investigación y diseñados para una finalidad puntual de investigación lingüística, como los siguientes:
 - *PaChEs* (Doval, Orduña y Wang, 2019) (en construcción)
 - *Corpus Paralelo de Español Inglés y Chino (CPEIC)* (Lu, 2013) (<http://140.116.245.146/cpeic/TriLin.html?role=teacher>)

- *RST Spanish-Chinese Treebank* (Cao, da Cunha y Iruskieta, 2018)
(<http://ixa2.si.ehu.es/rst/zh/index.php>)

Sin embargo, en la elaboración de la propuesta didáctica de nuestro trabajo, no vamos a utilizar todos los corpus mencionados arriba (once en total). Más en concreto, vamos a excluir el corpus *MyMemory*, los corpus del Grupo B, el corpus *PaChEs* y *RST Spanish-Chinese Treebank*. Esto es porque, en primer lugar, aunque *MyMemory* y *TAUS Data Cloud* son dos corpus similares creados a partir de las memorias de traducción para mejorar la calidad de traducción, presentan diferencias. Los textos recopilados en estos corpus proceden de las páginas web bi-/multilingües (sobre la web como corpus paralelo, véase el apartado 3.2.3) en distintos campos de especialidad. Sin embargo, cuando buscamos una misma palabra (p. ej. la palabra *pero*) en estos dos corpus, en la mayoría de ocasiones *MyMemory* presenta menos resultados (p. ej. 20 resultados aproximadamente) y frases alineadas más cortas (p. ej. *pero comencemos; pero no basta*, etc.). En cambio, *TAUS Data Cloud* nos ofrece más resultados (p. ej. 100 resultados) y frases alineadas más completas. Por eso, en este trabajo optamos por *TAUS Data Cloud* en vez de *MyMemory*. En segundo lugar, los corpus de Grupo B suelen servir como la base y la procedencia de los textos de los corpus de Grupo A, así que los datos incluidos en estos corpus son realmente accesibles con ayuda de las herramientas de búsqueda que ofrecen los corpus en Grupo A. Además, para consultar estos corpus del grupo B se requiere más conocimiento tanto lingüístico como técnico. Incluso en la página del *Corpus Paralelo de las Naciones Unidas* no se puede encontrar una interfaz para hacer consultas ni búsquedas de palabras. Así que es evidente que la utilización de estos dos corpus encierra mayor grado de dificultad. En tercer lugar, no vamos a utilizar el corpus *PaChEs* porque todavía está en construcción. A pesar de eso, cabe mencionar que el proyecto *PaChEs*, como uno de los primeros proyectos

centrados en crear un corpus paralelo chino-español especialmente para la enseñanza de chino y español como lengua extranjera, nos brindará muchas ideas sobre la aplicación didáctica de este tipo de recurso. En cuarto lugar, *RST Spanish-Chinese Treebank* es un corpus que recopila textos en español y traducciones en chino, creado por un grupo de investigación para favorecer los análisis comparativos discursivos entre chino y español (Cao, da Cunha y Irukieta, 2018). Sin embargo, la herramienta de búsqueda que ofrece este corpus solo nos presenta los resultados monolingües en vez de concordancias paralelas. Es decir, la búsqueda se limita a los dos subcorpus monolingües. Para acceder a los ejemplos alineados en dos lenguas los usuarios tienen que elegir entre las relaciones discursivas como justificar, concesión, causa, etc. (Ilustraciones 9 y 10). Esta particularidad, a nuestro entender, obstaculiza en gran medida la utilización de este recurso como corpus paralelo, dado que los alumnos no pueden hacer la búsqueda según sus propias necesidades para obtener resultados bilingües.

The screenshot shows the RST Spanish-Chinese Treebank interface. At the top is a dark red header with the title "RST SPANISH-CHINESE TREEBANK" in white. Below the header is a navigation bar with several tabs: HOME, RELATIONS, RELATIONS IN TREES, EDUS, SEARCH, SEARCH SPANISH, REFERENCES, and PRIVATE. The main content area displays three columns of relations, each with a title and a list of items. The first column is titled "PRESENTATIONAL RELATIONS" and includes items like "preparation", "background*", "Enablement and motivation", "enablement*", "motivation", "Evidence and justify", "evidence", "justify", "Anthithesis and concession", "anthithesis", "concession", "restatement and summary", "restatement*", and "summary*". The second column is titled "SUBJECT MATTER RELATIONS" and includes items like "elaboration", "means*", "circumstance", "solution-hood", "Conditional subgroup", "condition", "otherwise*", "unless", "unconditional*", "Ebaluation and interpretation", "interpretation", "evaluation", "Cause subgroup", "cause", "result" (highlighted with a red box), and "purpose". The third column is titled "MULTINUCLEAR" and includes items like "list", "disjunction", "joint*", "sequence", "contrast", "conjunction", "restatement-NN", and "same-unit".

PRESENTATIONAL RELATIONS	SUBJECT MATTER RELATIONS	MULTINUCLEAR
preparation	elaboration	list
background*	means*	disjunction
Enablement and motivation	circumstance	joint*
enablement*	solution-hood	sequence
motivation	Conditional subgroup	contrast
Evidence and justify	condition	conjunction
evidence	otherwise*	restatement-NN
justify	unless	
Anthithesis and concession	unconditional*	same-unit
anthithesis	Ebaluation and interpretation	
concession	interpretation	
restatement and summary	evaluation	
restatement*	Cause subgroup	
summary*	cause	
	result	
	purpose	

Ilustración 9 Consulta en *RST Spanish-Chinese Treebank* (1)

RST SPANISH-CHINESE TREEBANK

Other Treebanks
- Multiling
- Basque

HOME

RELATIONS

RELATIONS IN TREES

EDUS

SEARCH

SEARCH SPANISH

REFERENCES

PRIVATE

Search here:

All documents▼

search

No. of characters

empty for complete segment.

Relation: result% (51)

Left unit	Sense	Right unit	Relation type	Relation name	Document	Tagger	Area	Notes
除了教学用书, 我们的教室团队也设计安排了其他许多丰富多彩的课堂元素, 例如游戏、视频、歌曲等等。	<--	这些都使学生们能够在塞万提斯学院有效而愉快的学习到西班牙语。	result	N-S	BMCS_CHN2-GS.rs3	GS	BMCS_CHN	
A los libros de texto se suman multitud de actividades interesantes preparadas por el equipo de profesores, juegos, proyección de videos, canciones...	<--	todo ello hace que aprender español con el Instituto Cervantes sea no sólo eficaz sino muy divertido.	result	N-S	BMCS_ESP2-GS.rs3	GS	BMCS_ESP	
Todos los profesores son nativos y tienen experiencia como examinadores DELE en Shanghai.	<--	Por tanto, nadie mejor que ellos para ayudarte a aprobar tu examen.	result	N-S	BMCS_ESP4-GS.rs3	GS	BMCS_ESP	
具体而言, Mapfre北美市场的保费收入为21.03亿欧元, 而南美市场为18.23亿欧元。而去年情况相反, 北美收入为15.73亿欧元, 南美收入为20.95亿欧元。	<--	因此, 截止9月, 北美市场收入同比上涨33.7%, 而南美同比下滑13%。	result	N-S	CCICE_CHN1-GS.rs3	GS	CCICE_CHN	
另外, 财政部将在本周四再次回到市场拍卖中长期国债, 欲拍卖30亿至40亿欧元。	<--	因此, 财政部将在本周通过国债拍卖筹集70亿至90亿欧元资金。	result	N-S	CCICE_CHN3-GS.rs3	GS	CCICE_CHN	
西班牙BBVA银行将关闭葡萄牙分行26个营业点, 并解雇187名员工, 继续降低该国的营运成本。	<--	因此, BBVA将不断缩小葡萄牙的营业点, 直到最后剩下12个。	result	N-S	CCICE_CHN4-GS.rs3	GS	CCICE_CHN	

Ilustración 10 Consulta en *RST Spanish-Chinese Treebank* (2)

Por todo ello, en nuestro trabajo hemos decidido utilizar los cinco corpus paralelos chino-español fácilmente manejables en línea, es decir, *TAUS Data Cloud*, *Glosbe*, *Eshelper*, *ReversoContext*, y *CPEIC*. Y a continuación, vamos a presentar una serie de criterios con el objetivo de compararlos y extraer conclusiones sobre cómo y cuándo utilizarlos.

4.2 Criterios de análisis

En este trabajo fin de máster, tomamos la teoría de Buyse (2011, 2021) como referencia para comparar la posibilidad y la productividad de utilización en el aula de ELE de cada uno. Además, compararemos los corpus teniendo en cuenta las características y conceptos clave en el proceso de la construcción de los corpus, como se ha visto en el apartado 3.2.1 y 3.2.2. Es decir, analizaremos los corpus a partir de las siguientes preguntas:

- 1) Idiomas: ¿cuántos idiomas contiene este corpus? En cuanto al par de idioma chino-español, ¿qué variedades incluye? Como hemos mencionado en el apartado 3.1.4, los corpus que permiten limitar la búsqueda a ciertas áreas geográficas pueden

ayudar a los aprendices a aprender la diferencia de uso y significado en las diferentes variedades del español.

- 2) Utilidad: ¿con qué objetivo se ha construido este corpus? Generalmente, los corpus centrados en la enseñanza-aprendizaje de lenguas son más fáciles de manejar, recopilan más tipos de textos y proporcionan más anotación lingüística. En cambio, los corpus diseñados para las investigaciones de traducción suelen recopilar textos de especialidad y la calidad de textos y la precisión de alineación son relativamente baja.
- 3) Tamaño: ¿cuál es el tamaño de este corpus? y, ¿cuál es la procedencia de los textos recopilados? Cuantos más ejemplos de uso puedan ver los aprendices, más fácil será la observación y extracción de conclusiones.
- 4) Selección de textos (sobre la selección de los textos en la construcción de un corpus paralelo, véase el apartado 3.2.2): ¿los textos se han sometido a un proceso de selección y a una revisión manual? Los textos seleccionados por expertos y revisados manualmente siempre tienen mejor calidad y son más adecuados para presentar a los alumnos en el aula.
- 5) Lematización: ¿es un corpus lematizado? Como hemos dicho en el apartado 3.1.2.3, un corpus lematizado facilitará en gran medida los análisis a partir del corpus.
- 6) *Keyword* resaltado en ambas versiones de lengua: si la palabra que buscamos está resaltada tanto en el texto original como en la traducción, los aprendices pueden observar la correspondencia y equivalencia con más claridad.
- 7) POS-tagging⁴¹: ¿hay etiquetación gramatical? Esta información morfo-sintáctica ayudará a precisar la búsqueda. Por ejemplo, un corpus con la etiquetación

⁴¹ *Parts of speech*, es decir: categorías gramaticales (Buyse, 2011).

gramatical puede establecer la diferencia entre la forma sustantiva y la verbal de la palabra “vino”.

- 8) Definición: ¿se muestra la definición de las palabras que buscamos? Esta información semántica facilitará la comprensión y el aprendizaje de la palabra que buscamos.
- 9) Combinación de palabras: ¿se puede hacer búsqueda de combinación de palabras? Esta característica es importante especialmente para el aprendizaje de las colocaciones, locuciones, frases hechas, etc.
- 10) Búsqueda filtrada: ¿se pueden filtrar los resultados según el tema, el tipo de textos, etc.? Esta opción nos permite comparar los textos de diferentes géneros, registros, tipos de textos, etc. y nos facilitará los análisis comparativos entre diferentes textos.
- 11) Alineación: ¿cuál es el nivel de alineación de este corpus? Sobre los niveles de alineación, véase el apartado 3.2.2.
- 12) Información anotada: ¿hay información de los metadatos y el marcado textual? Como se ha visto en el apartado 3.2.2, este tipo de información nos permite recuperar y localizar la información relevante del corpus posteriormente.
- 13) Dirección de traducción (véase el apartado 3.2.2): ¿es un corpus unidireccional, bidireccional o multidireccional? Como hemos mencionado en el apartado 3.2.2, el equilibrio de la dirección de traducción parece fundamental para los alumnos a la hora de procesar los datos en el corpus.
- 14) Particularidad: ¿tiene algunas características particulares?

5. PRESENTACIÓN Y ANÁLISIS DE LOS CORPUS SELECCIONADOS

Una vez presentados los criterios que hemos seguido en este TFM para escoger los corpus, a continuación, presentamos los cinco corpus paralelos chino-español sobre los que trabajaremos, describimos las características de cada uno y los comparamos entre sí siguiendo los criterios que hemos establecido más arriba (objetivo 2), para saber cómo y cuándo se debe utilizar estos corpus en el aula de ELE⁴².

5.1 TAUS Data Cloud

TAUS Data Cloud (<https://data-app.taus.net/>) es un corpus paralelo creado a partir de memorias de traducción. Es una plataforma de repositorio para compartir, agrupar y aprovechar los datos de traducción. Contiene decenas de miles de millones de palabras en múltiples direcciones de traducción, 62 idiomas, 17 campos de especialidad bien definidos, 9 tipos de contenido procedentes de múltiples fuentes de datos, por ejemplo, memoria de traducción (TM) aportadas por miembros y usuarios de TAUS, datos públicos a disposición de las comunidades como grandes organizaciones e instituciones, etc.⁴³ Es una herramienta fácil de manejar. Para hacer la consulta en *TAUS Data Cloud*, debemos introducir la palabra y elegir los idiomas y seleccionar *Search*. Además, en este corpus, se permite limitar la búsqueda en función de diferentes criterios como campo de especialidad, tipo de texto, y propietario de datos.

⁴² Toda la información sobre los corpus se extrae de las páginas web de los corpus o estudios relacionados con la construcción de los corpus, disponible en: <https://www.taus.net/faq/topic/187> [Consulta: 15/06/2020].

⁴³ Información sobre *TAUS Data Cloud* extraída de la página <https://www.taus.net/data/data-for-machine-translation> y <https://www.taus.net/faq/topic/205>

me gusta

Spanish (International) > Chinese (PRC) Search

Less options

Industry: Select...
 Computer Hardware
 Computer Software
 Legal Services
 Leisure, Tourism, and Arts
 Medical Equipment and Supplies
 Pharmaceuticals and Biotechnology
 Professional and Business Services
 Stores and Retail Distribution
 Undefined Sector

Content Type:

Data Owner:

Source Segment		
Pero no me gusta el sentimentalismo.	但我不喜欢多愁善感。	
Eso no me importa, me gusta .	我并不介意，我喜欢这样做。	
En realidad no me gusta hablar de determinada tribu.	我确实不想谈起个别部落。	
No me gustan las actitudes de cowboy, pero tampoco me gusta decir: "Dejemos de trabajar".	我并不喜欢牛仔脾气，但我也不要说："我们不要开展任何活动"。	
Evidentemente, me gusta mi trabajo, pero quizá sea mejor que me detenga aquí.	但是我或许应该就此收场。	
A mí no me gusta hablar mucho; prefiero que haya más hechos.	我不喜欢说得太多；我希望有更多的行动。	
Me gusta ser Presidenta, así que esto quizá se prolongue un poco más de lo previsto.	我有点喜欢做主席，所以可能会要比预期的时间长一些。	

Ilustración 11 Búsqueda de la combinación léxica *me gusta* en *TAUS Data Cloud*

A continuación, vamos a describir las características del corpus *TAUS Data Cloud*, de acuerdo con los criterios presentados en el apartado XXXX:

- Es multilingüe. A la hora de hacer la búsqueda, se puede elegir entre español de España, México, América Latina, y español internacional. Por parte del chino, se distingue entre chino de República Popular China, Hongkong y Taiwan.
- La finalidad de este corpus es mejorar la calidad de traducción, servir de entrenamiento a sistemas de traducción automática y realizar investigaciones en el ámbito de tecnología.
- Contiene decenas de miles de millones de palabras en total⁴⁴, procedentes de 17 campos de especialidad bien definidos, 9 tipos de contenido procedentes de múltiples fuentes de datos, por ejemplo, memoria de traducción (TM) aportadas por miembros y usuarios de *TAUS*, datos públicos a disposición de las comunidades como grandes organizaciones e instituciones, etc.

⁴⁴ La información sobre el tamaño del corpus chino-español no está registrada.

- d) No hay selección de texto ni revisión manual.
- e) Es un corpus no lematizado.
- f) El *keyword* aparece resaltado en negrita solo en el texto original.
- g) No hay etiquetado gramatical (*POS-tagging*).
- h) No se muestra la definición de la palabra que buscamos.
- i) Es posible hacer búsqueda de combinaciones de palabras.
- j) Se puede limitar la búsqueda a determinados campos de especialidad (por ejemplo, servicio jurídico, ocio, turismo y arte, material médico, medicamento, biotecnología, etc.) y ciertos tipos de textos (por ejemplo, documentación financiera, instrucciones de uso, anuncio, reportaje, etc.).
- k) Alineado por frase (no KWIC).
- l) No hay información sobre el autor y/o traductor, título, fecha, información sobre la publicación, lengua original y la lengua de la versión
- m) Los textos recopilados son en múltiples direcciones de traducción, es decir, es un corpus multidireccional.

5.2 Glosbe

Glosbe (<https://es.glosbe.com/>) es un “corpuscionario” (véase el apartado 3.1.3) multilingüe en línea creado a partir de memorias de traducción. Su memoria de traducción proviene de traducciones hechas por humanos y la interfaz permite ver no sólo la traducción de la palabra, sino también cómo se comporta en la frase. Actualmente cuenta con 375 mil frases traducidas alineadas en el diccionario chino-español. *Glosbe* es un proyecto colaborativo y cada uno de sus usuarios puede añadir o eliminar traducciones para que los ejemplos en este diccionario sean reales⁴⁵. *Glose*

⁴⁵ Información sobre *Glosbe* extraída de la página <https://es.glosbe.com/zh/es>, <https://es.glosbe.com/es/zh> y <https://es.glosbe.com/partners>.

también es una herramienta fácil de usar. Comparado a *TAUS Data Cloud*, *Glosbe* nos proporciona más información sintáctico-semántica.



Ilustración 12 Búsqueda de la palabra *gustar* en *Glosbe* (1)



Ilustración 13 Búsqueda de la palabra *gustar* en *Glosbe* (2)

A continuación, describimos las características de *Glosbe*:

- a) Es multilingüe: incluye español, chino y unos dialectos de chino como Wu, Min Nan, etc.
- b) El objetivo principal del diccionario *Glosbe* es, en palabras de sus responsables, proporcionar el mejor diccionario multilingüe en línea, que incluya casi todos los idiomas existentes en el mundo.
- c) En el diccionario chino-español, cuenta con 200.149 frases traducidas. En el de español-chino, hay 175.809 frases traducidas. Es un diccionario en línea creado a partir de las memorias de traducción elaboradas por seres humanos. Las frases incluidas en este diccionario proceden de los documentos oficiales de grandes organizaciones internacionales en diferentes lenguas, las páginas web multilingües, etc. Además, *Glosbe* es un proyecto colaborativo y cada uno puede añadir (y eliminar) traducciones.
- d) No hay selección de texto ni revisión manual.
- e) Es un diccionario lematizado, pero en sentido muy limitado. Por ejemplo, si buscamos el infinitivo del verbo *gustar*, se nos presentan ejemplos en los que aparecen *gustar*, *gustará*, y *gustara* sin otras conjugaciones. Deducimos, por lo tanto, que se basa más en búsquedas por raíz que por lema.
- f) La palabra que buscamos está resaltada solo en la lengua original.
- g) Hay etiquetado gramatical (*POS-tagging*).
- h) Se muestra la definición de la palabra que buscamos.
- i) Se pueden hacer búsquedas de combinaciones de palabras.
- j) No se puede filtrar el resultado de la búsqueda según diferentes áreas temáticas, ni ciertos tipos de textos.
- k) Alineado por frase (no KWIC).
- l) No hay información sobre el autor y/o traductor, título, fecha, información sobre

la publicación, lengua original y la lengua de la versión.

- m) Es un corpus multidireccional.
- n) La particularidad de este corpus consiste en que se nos ofrece, en primer lugar, la lista de las consultas más comunes en el diccionario chino-español y español-chino, en segundo lugar, el audio de grabaciones de la palabra y, en tercer lugar, todo el paradigma de la conjugación de la palabra en caso de ser verbo.

5.3 Xiyuzhushou (Eshelper)

Xi Yu Zhu Shou (<https://www.esdict.cn/>) es uno de los diccionarios bilingües chino-español en línea más usados entre los alumnos chinos. Es un “corpuscionario” (véase el apartado 3.1.3) diseñado especialmente para los aprendices sinohablantes de español para consultar palabras, conjugaciones de verbo, traducciones, ejemplos de uso, etc. Recopila 450.000 bisegmentos alineados basando los datos de la web (sobre la web como corpus paralelo, véase el apartado 3.2.3) como los textos en las noticias, la Wikipedia, los libros, los manuales, las pruebas y los exámenes, etc.⁴⁶ Además, se permite acceder al diccionario bilingüe de términos jurídicos, económicos y de negocio⁴⁷. Puesto que es un recurso diseñado especialmente para los aprendices sinohablantes de español, las frases recopiladas parecen más sencillas, cotidianas y fáciles de comprender.

⁴⁶ Información sobre *Eshelper* extraída de la página <http://www.eudic.net/v4/es/app/eshelper>.

⁴⁷ *XiBanYaYuFaLuWaiMaoJingMaoCiHui* (Diccionario del Léxico español jurídico, económico y de negocio), disponible en *Eshelper*.

gustar

常用词 专八 A1A2

音标: [guʃ'tar] 发音: 全球发音: 生词本: 添加笔记:

西汉-汉西词典 西语动词变位 西语例句库 近反义词 西西词典 西语维基词典 西英词典 全文检索

有奖纠错 划词

一 西汉-汉西词典

提示: 点击查看 [gustar 的动词变位](#)

tr.

1. 品味, 品尝.

Gustó el vino y le pareció excelente. 他尝了一下酒, 觉得味很醇厚。

2. 体验, 领略, 经历.

Gustaron las mieles del triunfo. 他们已领略到胜利的甜蜜。

l→ intr.

1. «de»

喜欢, 喜爱:

Gusta de conversar con la gente. 他喜欢和人交谈。

Gusto de leer. 我喜欢读书。



赞 踩 改进

Ilustración 14 Búsqueda de la palabra *gustar* en *Eshelper* (1)

gustar

常用词 专八 A1A2

音标: [guʃ'tar] 发音: 全球发音: 生词本: 添加笔记:

西汉-汉西词典 西语动词变位 西语例句库 近反义词 西西词典 西语维基词典 西英词典 全文检索

有奖纠错 | 划词

— 原声例句





1. Me gusta salir con mis amigos. 
我喜欢和朋友出去玩。
《视频版初级西语课程》
2. Pues...me gusta bastante la música disco, las motos me encantan, me gusta bailar... 
唔...我挺喜欢迪斯科音乐的, 摩托车也使我着迷, 我还喜欢跳舞...
《NUEVO VEN 1》
3. Me gustaría hacerle unas preguntas relacionadas con la cocina, ¿vale? A ver, ¿le gusta cocinar? 
我想问您一些关于饮食的问题, 可以吗? 开始吧, 您喜欢做饭吗?
《Dele B1 Expresión e interacción orales》
4. Te las vamos a leer y nos gustaría saber qué piensas tú. 
我们将念给你听, 然后听听你的想法。
《Aula Internacional 2》



Ilustración 15 Búsqueda de la palabra *gustar* en *Eshelper* (2)

A continuación, presentamos las características del corpus paralelo que ofrece *Eshelper*:

- a) Es bilingüe.
- b) *Eshelper* es un “corpuscionario” bilingüe diseñado como una herramienta para los aprendices sinohablantes que aprenden español.
- c) Los ejemplos de uso provienen de los recursos abiertos al público en la web. Recopila 165 mil ejemplos extraídos por los materiales didácticos, los exámenes, los libros, las películas, etc., lo cual implica la facilidad de comprensión de los ejemplos.
- d) No hay selección de texto. Algunos de los ejemplos están revisados manualmente.
- e) Es lematizado.
- f) La palabra que buscamos aparece resaltada en diferente color en el texto original.
- g) Hay etiquetado gramatical (*POS-tagging*).
- h) Incluye la definición de la palabra que buscamos, a veces acompañada de fotos o ilustraciones.

- i) Se permite la búsqueda de combinaciones de palabras.
- j) No se puede filtrar el resultado de la búsqueda según diferentes áreas temáticas, ni ciertos tipos de textos. Pero se puede filtrar los resultados por la categoría gramatical.
- k) Alineado por frase (no KWIC).
- l) No hay información sobre el autor y/o traductor, fecha, información sobre la publicación, lengua original y la lengua de la versión, pero hay información sobre el título de documento.
- m) Es un corpus multidireccional, puesto que incluyen textos cuya versión original es no solo en español y chino, sino también en otras lenguas (por ejemplo, japonés, inglés, etc.)
- n) Al ser un “corpuscionario” especialmente diseñado para los alumnos sinohablantes de español, nos presenta más información en la página de resultados de consulta, por ejemplo, los sinónimos y antónimos, las conjugaciones del verbo, las unidades fraseológicas que incluyen esta palabra buscada, los audios de la pronunciación, las entradas de la palabra en el diccionario de español monolingüe, la información de la palabra en la Wikipedia, la lista de las consultas más comunes en el diccionario, etc.

5.4 ReversoContext

ReversoContext (<https://context.reverso.net/traduccion/>) es un “corpuscionario” que utiliza millones de textos bilingües, procesados por complejos algoritmos elaborados con inteligencia artificial. Incluyen más de 100 pares de idiomas y miles de millones de traducciones extraídas de diálogos de películas, documentos oficiales, páginas web, periódicos. Es decir, los ejemplos se extraen de contextos de la vida real y cubren una

gran variedad de registros de habla⁴⁸. Aparte de los ejemplos alineados, los usuarios pueden acceder a otra información (p. ej. sinónimos, definición, imágenes, entrada en Wikipedia, etc.) por medio de hacer clic en los enlaces correspondientes⁴⁹. Cabe mencionar que, en comparación con los dos “corpuscionarios” mencionados arriba, una particularidad de ReversoContext consiste en que se permite ver el contexto más amplio de una traducción y verificar la fuente de un ejemplo. Como hemos dicho en el apartado 3.2.2, esta recuperación de información puede ser relevante para los usuarios cuando tienen dudas sobre los ejemplos.

ReversoContext TRADUCCIÓN | DICcionario | CORRECTOR | CONJUGACIÓN | SINÓNIMOS | DOCUMENTS new

gustar Español Chino BETA

Conjugar este verbo

Sugerencias: va a gustar gustar mucho

Traducción de "gustar" en chino

Verbo

喜欢 尝试 爱 喜爱 尝 享受 讨好 尝味 关心 品尝 担心 爱好 会爱听 尝 会很不高兴

No pensé Waller podría gustar a nadie. 没想到沃勒也会喜欢人

¿No te debería gustar tu nombre? 大概? 你的名字自己得喜欢啊

Lo importante no es gustar, sino disgustar. 重要的事是不喜欢 这使我很高兴

Preferiría más gustar que ser perfecta. 我宁愿被人喜欢 而不是做个完人

Depende. ¿Cuánto me tiene que gustar? 看情形啦, 我要多喜欢他才好?

Ilustración 16 Búsqueda de la palabra *gustar* en ReversoContext

⁴⁸ Información sobre ReversoContext extraída de la página <https://context.reverso.net/traduccion/about>.

⁴⁹ Los enlaces dirigen a los sitios web como ReversoSynonyms, ReversoDiccionario, Google, Wikipedia, etc.

Las características de *ReversoContext* son las siguientes:

- a) Es multilingüe. Contiene más de 100 pares de lenguas, que incluyen chino-español.
- b) La finalidad de *ReversoContext*, en palabras de sus responsables, es establecer una herramienta de traducción y diccionario en línea basado en la inteligencia artificial.
- c) Contiene miles de millones de traducciones. Pero la información sobre el tamaño del corpus chino-español no está registrada. Los ejemplos se extraen de diálogos de películas, documentos oficiales, páginas web, periódicos. Es decir, los ejemplos cubren una gran variedad de registros de habla.
- d) No hay selección de textos ni revisión manual.
- e) Es un corpus no lematizado.
- f) La palabra que buscamos (*keyword*) aparece resaltada en las dos lenguas.
- g) Hay etiquetado gramatical (*POS-tagging*).
- h) Para acceder a la definición de la palabra que buscamos, hay que hacer clic a un enlace que dirige a *ReversoDiccionario*.
- i) Se permite buscar combinaciones de palabras.
- j) No se puede limitar la búsqueda a diferentes tipos de textos, ni a determinadas áreas temáticas. Pero se permite filtrar los resultados por las categorías gramaticales.
- k) Alineado por frase (no KWIC).
- l) No hay información sobre el autor y/o traductor, título, fecha, información sobre la publicación, lengua original y la lengua de la versión, pero hay enlaces que conducen a la fuente de los textos.
- m) Es multidireccional.
- n) Se permite recuperar un contexto más amplio de las traducciones y también es posible reproducir la pronunciación de cada frase.

5.5 CPEIC

El *Corpus Paralelo de Español, Inglés y Chino* (CPEIC) (<http://140.116.245.146/cpeic/TriLin.html?role=teacher>) fue creado por un grupo de investigación de la National Cheng Kung University con el objetivo de proporcionar datos fiables para facilitar el análisis contrastivo de tres lenguas: español, inglés y chino; que son respectivamente la lengua meta (L3), la primera lengua extranjera (L2) y la lengua materna (L1) de los estudiantes sinohablantes. Desde 2007 hasta 2017, se ha recogido una gran cantidad de materiales idiomáticos (*p. ej.* textos en la Biblia, cuentos y documentos de Naciones Unidas, subtítulos de películas etc.), y se continúa mejorando la construcción del corpus. Es un corpus trilingüe disponible en línea que incluye tanto el lenguaje escrito como el lenguaje oral.⁵⁰ A diferencia de los corpus mencionados más arriba, los textos en *CPEIC* son seleccionados por expertos. Además, este corpus permite la búsqueda en español, inglés y chino, y nos presenta frases alineadas y paralelas en estas tres lenguas. Asimismo, la búsqueda puede ser filtrada por la categoría gramatical (POS) y los temas de textos.

The screenshot displays the CPEIC search interface. On the left, there are filters for 'Temas' (Topics) and 'Idioma' (Language). The 'Temas' filter is expanded, showing 'Cuentos' (0), 'NU: Escrito' (1), 'NU: Oral' (1), 'Biblia' (1), and 'Películas' (1). The 'Idioma' filter is set to 'Español'. The search term 'gustar' is entered, and the POS is set to '(todas)'. The 'Filtro' is set to 'Obligatorio'. The search results are displayed in a table with columns for 'Información', 'Artículo 1', and 'Artículo 2'. The 'Información' row shows 'Artículos: 1', 'Oraciones: 1', and 'Palabras: gustar(1)'. The 'Artículo 1' row shows the title 'Hable con ella# Talk to Her # 悄悄告诉她' and the result 'Resultado #1 Oración #911'. The 'Artículo 2' row shows the aligned sentences in Spanish, English, and Chinese.

Información:	Artículo 1:	Artículo 2:
Artículos: 1 Oraciones: 1 Palabras: gustar(1)	Título: Hable con ella# Talk to Her # 悄悄告诉她 Resultado #1 Oración #911	es: Y claro que me gusta , ¿ cómo no me va a gustar ? in: Of course I like her . ch: 怎麼會不喜歡?

Ilustración 17 Búsqueda de la palabra *gustar* en *CPEIC*

⁵⁰ Información sobre *CPEIC* extraída de la “Introducción” de la página <http://140.116.245.146/cpeic/TriLin.html?role=teacher>.

A continuación, veremos las características de este corpus:

- a) Es trilingüe, que incluye español, inglés y chino tradicional. Y las concordancias paralelas son frases alineadas en tres lenguas (no KWIC).
- b) El *CPEIC* tiene como propósito favorecer el desarrollo de los análisis contrastivos de tres lenguas paralelas: español, inglés y chino, con lo que se puede facilitar la enseñanza de español para los alumnos sinohablantes.
- c) Hasta la fecha, este corpus incluye 1.217.971 y 1.190.081 palabras en español e inglés y 1.543.580 caracteres en chino. La recolección y procesamiento de los materiales del corpus se divide en cuatro partes: Corpus de La Biblia (lenguaje escrito), Corpus de cuentos (lenguaje escrito), corpus de Naciones Unidas (lenguaje escrito y oral) y Corpus de películas (lenguaje escrito y oral).
- d) Hay selección de texto por expertos y revisión manual.
- e) Es un corpus no lematizado.
- f) Las palabras clave aparecen en diferente color en las frases tanto en español como en inglés y chino.
- g) Hay etiquetado gramatical (*POS-tagging*).
- h) No se muestra la definición de la palabra que buscamos.
- i) No se pueden hacer búsquedas de combinaciones de palabras.
- j) Se puede limitar la búsqueda a determinados temas en diferentes subcorpus (por ejemplo, cuentos, corpus de Naciones Unidas, Biblia y Películas.) y a ciertos tipos de textos (por ejemplo, oral y escrito), así como comparar los resultados entre diferentes subcorpus. Asimismo, se puede buscar una palabra limitando la categoría gramatical.
- k) Alineado por frase (no KWIC).
- l) No hay información sobre el autor y/o traductor, fecha, información sobre la publicación, lengua original y la lengua de la versión, pero hay información sobre

el título de documento.

m) Es un corpus multidireccional.

Para tener una visión general, a continuación, mostramos en la siguiente tabla la comparación de estos cinco corpus de acuerdo con los criterios mencionados en el apartado 4.2⁵¹, con la que se puede observar que cada corpus tiene sus puntos a favor y en contra.

⁵¹ En la tabla vamos a excluir el criterio del tamaño porque, por un lado, la información sobre el tamaño del corpus chino-español no están registrada en la página de *TAUS Data Cloud* y *ReversoContext* y, por otro lado, en los otros tres corpus los editores utilizan diferentes parámetros para describir el tamaño y somos conscientes de que estos parámetros no son comparables.

Corpus Criterios	TAUS	Glosbe	Reverso Context	Eshelper	CPEIC
1. Multilingüe	+	+	+	-	+
2. Finalidad ⁵²	-	+	+	+	+
3. Selección y revisión de los textos	-	-	-	-	+
4. Lematizado	-	+	-	+	-
5. <i>Keyword</i> destacado en dos versiones de lengua	-	-	+	-	+
6. Etiquetado gramatical (POS-tagging)	-	+	+	+	+
7. Etiquetado semántico (definiciones)	-	+	+	+	-
8. Combinaciones de palabras	+	+	+	+	-
9. Búsqueda filtrada por tema, tipo de texto, o categoría gramatical	+ / + / -	- / - / -	- / - / +	- / - / +	+ / + / +
10. Alineado por frase	+	+	+	+	+
11. Los metadatos y el marcado textual	-	Fuente	Fuente	Título	Título
12. Traducción multidireccional	+	+	+	+	+
13. Otras funciones	-	+	+	+	-

Tabla 1 Comparación de Taus Data Cloud, Glosbe, ReversoContext, Eshelper y CPEIC. Elaboración

propia

⁵² Diseñado para facilitar la enseñanza-aprendizaje de lenguas

5.6 Análisis de los corpus paralelos chino-español seleccionados

Tras describir las características de los cinco corpus disponibles para elaborar nuestra propuesta didáctica, hemos notado que existen ciertas deficiencias y problemas de los corpus a la hora de usarlos en el aula. En primer lugar, consideramos la calidad y la cantidad de textos en los corpus un gran problema. Entre los cinco corpus seleccionados, cuatro⁵³ de ellos toman la web como corpus (véase el apartado 3.2.3); es decir, los textos incluidos en estos corpus no han sido seleccionados ni revisados manualmente por expertos. Además, el proceso automático de construcción de un corpus a veces provoca que las frases no estén bien escritas ni bien alineadas y no haya correspondencia entre las dos versiones de lengua, puesto que a menudo en el proceso de la alineación por máquina se producen varias incidencias. Por ejemplo, como muestra la Ilustración 18, la frase en español es *tengo que hablar contigo*, mientras que la frase en chino es 我? 爱 死了 ,? 爱 死了- 我 必? 须 跟 你? 谈 (Yo? amo, ?amo, tengo que hablar contigo). Obviamente ha surgido un error tanto en la forma como en la alineación en este caso.

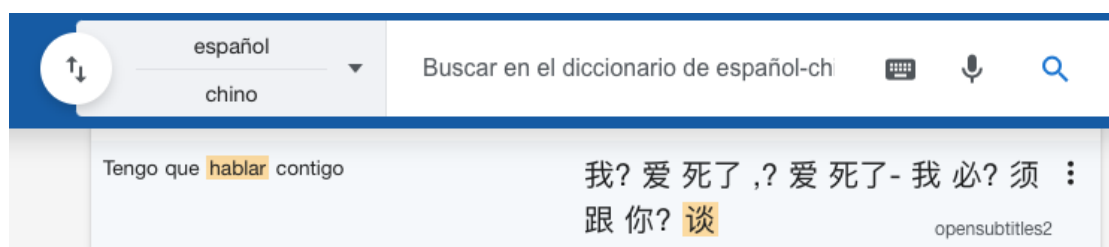


Ilustración 18 Búsqueda de la palabra *hablar* en *Glosbe*

Entre los cinco corpus analizados, el *CPEIC* nos proporciona textos de mejor calidad, ya que es un corpus construido por un grupo de investigación de expertos. Además, en

⁵³ Taus Data Cloud, Glosbe, ReversoContext y Eshelper.

comparación con otros corpus, a pesar de basarse también en los recursos de la web, *Eshelper* recopila muchos textos y ejemplos procedentes de materiales didácticos, obras publicadas y exámenes oficiales, lo cual asegura en gran medida la calidad de los ejemplos. Cabe mencionar que el corpus *PaChEs*, aunque todavía está en construcción, ha recopilado textos estrictamente seleccionados y la alineación de las frases ha sido bien revisada manualmente. Así que se considera como una herramienta muy productiva para favorecer la enseñanza-aprendizaje de español en el futuro estudio en este ámbito. Por otro lado, aunque los datos del *CPEIC* resultan más fiables, el tamaño de este corpus es muy limitado, así que en las búsquedas salen escasas concordancias. De ahí que se obstaculice la observación de los alumnos y la comparación del uso de las palabras clave entre diferentes situaciones textuales. En comparación con *CPEIC*, los demás corpus que toman los datos de la web tienen relativamente mayor tamaño, pero también los consideramos limitados. En definitiva, observamos que, como hemos visto en el apartado 3.2.3, los corpus basados en los datos extraídos de la web tienen mayor tamaño pero peor calidad de texto. En cambio, los corpus creados de una forma más manual tienen un tamaño más limitado pero mejor calidad de texto.

En segundo lugar, se observa que entre los cinco corpus seleccionados solo *Glosbe* y *Eshelper* son lematizados, y el nivel de lematización de *Glosbe* es muy limitado. Sin duda alguna, la ausencia de lematización limitará su utilidad en el aprendizaje-enseñanza de lenguas. Por ejemplo, en el caso del aprendizaje de una colocación que incluya un verbo, un corpus que permite la búsqueda por lemas posibilitará que los alumnos dispongan de más ejemplos para observar la combinación más habitual y extraer sus propias conclusiones.

En tercer lugar, la información de anotación (la anotación lingüística, los metadatos, el marcado textual), se encuentra muy limitada en todos estos cinco corpus. Por una parte, la anotación lingüística limitada restringe el nivel de análisis al que se pueda someter un corpus. Por otra parte, la ausencia de los metadatos sobre la obra, el autor, la fecha, etc. impide la recuperación de información relevante por parte de los usuarios cuando tienen dudas. Y como bien dice Frankenberg-Garcia (2005), a la hora de emplear los corpus paralelos en el aula, es importante tener en cuenta que las concordancias paralelas contienen no solo dos lenguas, sino también dos tipos de textos, que son el texto original y la traducción. Así que la falta de la información sobre la dirección de traducción de los textos puede provocar problemas para los alumnos al procesar los datos.

En cuarto lugar, otra deficiencia consiste en la imposibilidad de comparar entre géneros, temas, países, etc. en la mayoría de estos corpus. Esto se debe, por una parte, a que los textos que se recogen en estos corpus muestran poca diversidad en cuanto al género, medio, tema, área geográfica, etc. Por otra parte, entre estos cinco corpus, solo hay dos (*TAUS Data Cloud* y *CPEIC*) que nos permite seleccionar el tema que deseemos buscar. Es decir, la mayoría de los corpus no disponen de sistema para realizar una búsqueda filtrada. Esa debilidad limitará en gran medida el uso de los corpus paralelos en el aula de EFE (Español con Fines Específicos).

Teniendo en cuenta todo lo expuesto, consideramos que el que más se acerca al “corpus paralelo ideal” para utilizar en el aula de español sería el de *Eshelper*, considerando que es una herramienta especialmente diseñada para los alumnos sinohablantes de español, y la calidad de los textos es relativamente mejor. Sin embargo, cabe destacar que cada corpus tiene sus pros y contras. Así que es recomendable utilizar estos corpus de forma

combinada, con el objetivo de tener más ejemplos variados. Además, consideramos que la utilización de los corpus basados en los datos de la web⁵⁴ por parte de los alumnos de nivel inicial requiere la instrucción o supervisión del docente, ya que las frases en estos corpus no están siempre bien redactadas y traducidas. Por supuesto, los cinco corpus pueden ser de ayuda para los alumnos de nivel avanzado para hacer consultas autónomas (de tipo “self-access”) durante las tareas de escritura, puesto que los alumnos de este nivel ya son capaces de distinguir los pares de textos que no tienen correspondencia y escoger los ejemplos que les parecen útiles. Asimismo, los corpus que cuentan con un sistema para filtrar los resultados de búsqueda (*TAUS Data Cloud* y *CPEIC*) puede ser una buena herramienta a la hora de comparar un fenómeno lingüístico entre diferentes tipos de textos o áreas temáticas.

⁵⁴ Taus Data Cloud, Glosbe, ReversoContext y Eshelper.

6. PROPUESTA DE ACTIVIDADES

6.1 Introducción

En este apartado, vamos a presentar algunas posibilidades y ejemplos del uso de los cinco corpus paralelos chino-español seleccionados para los alumnos sinohablantes siguiendo los usos didácticos que hemos mencionado en los apartados anteriores. Nos interesan especialmente los usos directos y los usos como herramienta de consulta de tipo *self-access*. En nuestras propuestas vamos a presentar unas series de actividades dirigidas a los alumnos, acompañadas de nuestra propia reflexión teórica. Antes de presentar las propuestas, cabe aclarar unas pautas a la hora de trabajar con los cinco corpus paralelos en el aula.

En primer lugar, como hemos dicho en el apartado 5.6, una de las principales deficiencias de los corpus paralelos chino-español disponibles en línea consiste en la baja calidad de los textos. Así que, en nuestra propuesta, consideramos oportuno incluir dos diferentes maneras de utilización de los corpus paralelos: por una parte, para los alumnos de nivel inicial, es recomendable utilizar las concordancias paralelas extraídas y revisadas por el docente en el aula, puesto que, como hemos mencionado en apartado 5.6, los datos en estos cinco corpus paralelos no son siempre fiables. Por otra parte, se insta a los alumnos de nivel intermedio o avanzado a que hagan consultas y búsquedas ellos mismos en los corpus paralelos para obtener las frases paralelas, para observar los ejemplos y para extraer sus propias conclusiones. En este sentido, conviene especificar que en *Eshelper*, una herramienta diseñada especialmente para el aprendizaje de español, la calidad de los textos es relativamente mejor; por eso, consideramos que el nivel de los alumnos no limitará el uso de este corpus en el aula.

En segundo lugar, como hemos dicho en el apartado 3.1.4, los corpus permiten a los alumnos observar los casos repetidamente, verificar sus hipótesis y extraer sus conclusiones de modo inductivo. A ese respecto, cuantos más ejemplos se puedan obtener, más fácil será el proceso de extraer su propia conclusión. Por todo ello, para los alumnos de nivel intermedio y alto, es recomendable hacer la búsqueda de una misma palabra o combinación léxica en los cinco corpus, es decir, usar los corpus de forma combinada con el objetivo de obtener más ejemplos.

En tercer lugar, en el caso de que los alumnos tengan que hacer búsquedas ellos mismos, se requieren la conexión de Internet y el acceso a un dispositivo como tableta, ordenador, o móvil⁵⁵. Mientras que cuando se usan directamente las concordancias paralelas extraídas por el docente en el aula, los profesores pueden imprimir las concordancias en papel o mostrarlas con la ayuda de los proyectores.

Teniendo en cuenta todo lo mencionado, a continuación, vamos a presentar nuestras propuestas de actividades en la enseñanza de locuciones, marcadores del discurso y otras posibilidades utilizando los cinco corpus paralelos chino-español que hemos seleccionado.

6.2 Actividades para aprender locuciones

Una de las propuestas que lanzamos en este trabajo va orientada al aprendizaje de las locuciones con los corpus paralelos. Según el *Diccionario de la Lengua Española* (DLE, 23ª edición), las locuciones son “grupo de palabras que funcionan como una sola pieza léxica con un sentido unitario y cierto grado de fijación formal”. Cabe señalar que las

⁵⁵ Entre los cinco corpus seleccionados, ReversoContext y Eshelper tienen sus aplicaciones disponibles gratuitamente para móviles.

locuciones forman una parte de las unidades fraseológicas, las cuales constituyen un claro obstáculo para los aprendices sinohablantes de español debido a su idiomatidad, fijación, sentido cultural, etc. (Wu, 2016). Incluso en muchas ocasiones las locuciones no tienen equivalente en chino. Por todo ello, no es de extrañar que el aprendizaje de las locuciones constituya una gran dificultad en el aprendizaje del ELE para los estudiantes sinohablantes (Wu, 2016)⁵⁶. Conviene mencionar que, teniendo en cuenta la extensión de este trabajo, no nos vamos a detener aquí en revisar el marco teórico relacionado con locuciones porque el objetivo de este trabajo consiste en conocer las aplicaciones didácticas de los corpus paralelos (objetivo 1) y utilizarlos para presentar unas propuestas en el aula de ELE (objetivo 3).

No obstante, a nuestro juicio, los corpus paralelos chino-español pueden ser de gran utilidad para elaborar un amplio repertorio de ejercicios en el aprendizaje de locuciones para los estudiantes sinohablantes. Por ejemplo, se pueden utilizar para, por un lado, ayudar a los aprendices a inducir los significados idiomáticos de las locuciones y, por otro lado, comprobar si el alumno ha captado sus significados y usos. Evidentemente, la utilización de los corpus paralelos favorece el aprendizaje de las locuciones de manera contextualizada. Es decir, de esta forma, los estudiantes pueden aprender las locuciones a nivel textual, comunicativo y pragmático. A continuación, vamos a proponer una serie de ejercicios que se pueden explotar con los corpus paralelos en la enseñanza de las locuciones, dirigidos a los alumnos sinohablantes de nivel intermedio, con el objetivo de facilitar la comprensión y memorización de las locuciones y que los alumnos comprendan los significados idiomáticos de las locuciones y sepan cómo utilizarlas en la vida real.

⁵⁶ Sobre la fraseología española, puede verse Corpas Pastor (1996); Ruiz Gurillo (1997); Penadés (1999). Sobre la enseñanza de las locuciones para los estudiantes sinohablantes, puede verse Penas Ibáñez y Zhang (2012); Liu (2015); Wu (2016).

ACTIVIDAD 1. Y aunque todos hicieron cosas realmente encantadoras, aquella vez Carla Simpatías **dejó a todos con la boca abierta**.

尽管所有人都做的很好，但是这次 Carla Simpatías 着实让大家**目瞪口呆**。

(Fuente: Eshelper y elaboración propia)

1. En este ejemplo arriba, has visto que en español hay muchas expresiones que tienen un significado no literal. Observa las siguientes oraciones que incluyen la palabra *boca*, deduce qué significan realmente las expresiones destacadas e intenta explicarlo en español (en pareja). ¿Reconoce algunas expresiones similares en chino para expresar el mismo significado?

Se me abría la boca por haber dormido poco.
Cuando ve los pasteles, se le hace la boca agua .
No se le cae de la boca tu nombre .
La noticia corrió de boca en boca y ya lo saben todo.
Andar por ese lugar es meterse en la boca del lobo .
Te fuiste de la boca y ahora lo sabe todo.

(Fuente: Eshelper y elaboración propia)

Antes que nada, en la actividad 1 utilizamos un par de frases paralelas extraída de *Eshelper* como un ejemplo para introducir nuestro tema. En este ejemplo, la locución *dejar a alguien con la boca abierta* y su equivalente en chino (目瞪口呆 [ojo, mirar fijamente, boca, abierta y sin palabras]), se constituyen por los mismos elementos lingüísticos y comparten la misma base metafórica y el mismo significado idiomático (según DLE, suspenso o admirado de algo que se ve o se oye). Por lo tanto, es una

locución fácil de comprender y memorizar para los alumnos sinohablantes una vez presentadas simultáneamente estas frases paralelas. Este ejemplo ayudará a los alumnos a comprender qué es el significado idiomático, lo cual es diferente del significado literal en este caso. Con la ayuda de la correspondencia que muestran las frases paralelas, como hemos dicho en 1.2.4, los alumnos pueden alinear los conceptos mentalmente y familiarizarse con las estructuras lingüísticas. En la actividad 1, cuya duración debería ser de unos 20 minutos, seleccionamos algunas locuciones que contienen los constituyentes iguales o bien comparten las mismas bases metafóricas y, en esta etapa, solo presentamos los textos originales (en español) en los que aparecen las locuciones para que los alumnos deduzcan los significados idiomáticos de modo inductivo gracias a la observación de los contextos y basarse en su propio conocimiento previo de su lengua materna, en nuestro caso, el chino. En esta propuesta, todas las concordancias paralelas se consideran como input real, fiable y comprensible, como hemos dicho en el apartado 3.1.4.

ACTIVIDAD 2. Estas expresiones las llamamos *locuciones*. Ahora, observa estos ejemplos y las traducciones, y busca en las oraciones en chino los equivalentes de las locuciones destacadas en español. ¿Es la misma expresión que se te ha ocurrido en la actividad anterior? Asocia, después, las locuciones con sus equivalentes en chino.

Se me abría la boca por haber dormido poco.

因为睡眠时间少，我老打哈欠。

Cuando ve los pasteles, **se le hace la boca agua**.

他看到蛋糕的时候，就馋的流口水。

No **se le cae de la boca tu nombre**.

你的名字他老是挂在嘴边。

La noticia **corrió de boca en boca** y ya lo saben todo.

那消息已口口相传，现在大家都已知道此事

Andar por ese lugar es **meterse en la boca del lobo**.

在那地方行走，等于去冒无谓的风险。

Te fuiste de la boca y ahora lo sabe todo.

是你说漏了嘴，现在他全知道了。

(Fuente: Eshelper y elaboración propia)

Estar/quedar/dejar la boca abierta	说漏嘴
Abrirse la boca	身处险境
Hacerse la boca agua a uno	口口相传
Andar/correr/ir de boca en boca	流口水
Meterse en la boca del lobo	打呵欠
Irse de boca/ irse la boca a uno	目瞪口呆

En la actividad 2 (duración: 10 minutos aproximadamente), los alumnos están expuestos a las concordancias paralelas extraídas del corpus *Eshelper* que les ayudarán a establecer la correspondencia entre dos lenguas. Esta correspondencia facilitará la comprensión y memorización de las locuciones para los alumnos sinohablantes, puesto que la mayoría de estas locuciones comparten el elemento *boca* (口/嘴) y las otras tampoco son difíciles de comprender con la ayuda del contexto y traducción. Después, la actividad para asociar las locuciones con sus equivalentes obliga a los alumnos a repasar estas locuciones, lo cual favorecerá la retención de estas unidades léxicas. A nuestro juicio, esta forma contextualizada y asociada a conocimientos previos de los alumnos favorecerá al aprendizaje de estas combinaciones léxicas.

ACTIVIDAD 3. Has visto que, en muchas ocasiones, las expresiones en español y en chino que tienen el mismo significado contienen un par de constituyentes iguales (por ejemplo, *boca* [口/嘴]). En otras ocasiones, aunque no contienen ningún elemento igual, puedes deducir qué quiere decir según su significado literal y el contexto. Ahora, observa las siguientes oraciones y las traducciones y escribe una palabra que consideras adecuada en cada hueco. Ojo: las palabras serían las de partes de cuerpo.

He comprado un coche barato de segunda _____.

我买了一辆价钱便宜的二手车。

¡Válgate Dios, mujer, qué de cosas has ensartado unas en

otras, sin tener _____ ni _____!

上帝保佑，老伴儿，你干吗要把这些没头没尾的事连在一起？

La montaña no está lejos, en un abrir y cerrar de _____ ya llegamos.

山不太远，一眨眼的功夫我们就到了。

Eso no me parece estudiar, ni quemar las _____.

这看起来可不像学习，或者啃书。

Él trata siempre de echar una _____ a los verdaderamente necesitados.

他总试图向需要的人伸出援手。

¡No me sigas tomando el _____, que no sabes cómo me duele!

别再继续耍我了，你不知道我有多失望!

(Fuente: Eshelper, ReversoContext y elaboración propia)

En la actividad 3, cuya duración sería de unos 10 minutos, se ha reflejado con claridad la ventaja del uso de los corpus paralelos (*Eshelper*) en el aula de ELE. Entre estas locuciones, en esta actividad, muchas muestran la equivalencia total en dos lenguas, por ejemplo, *segunda mano* y 二手 (segunda, mano); *en un abrir y cerrar de ojos* y 一眨眼的功夫 (en un abrir y cerrar de ojos); *echar una mano* y 伸出援手 (echar, mano para ayudar). Algunos muestran la equivalencia parcial como *no tener pies ni cabeza* y 没头没尾 (sin, cabeza, sin, rabo). También hay casos con equivalencia nula como *quemarse las pestañas* y 啃书 (comer, libro). Las traducciones en el caso de equivalencia total y parcial ayudarán a los alumnos a rellenar los huecos y memorizar las locuciones con más facilidad. Además, cuando las locuciones no tienen equivalencia entre dos lenguas, las frases paralelas facilitarán la comprensión de los significados idiomáticos.

ACTIVIDAD 4. Busca en *Eshelper*, *TAUS Data Cloud*, *Glosbe*, *ReversoContext*, y *CPEIC* una palabra (una parte de cuerpo como *boca*, *mano*, *cabeza*...) y escoge tres ejemplos que contengan las locuciones en español que te resulten más interesantes. Compártelas con tus compañeros.

ACTIVIDAD 5. ¿Podrías redactar un pequeño texto utilizando las locuciones que has aprendido en estas actividades? En caso de que tenga dudas, consulta las cinco herramientas mencionadas arriba a ver si te sirve de ayuda.

En estas las actividades 4 y 5, que duran unos 30 minutos, se estimula a los alumnos a hacer consultas ellos mismos en los cinco corpus paralelos chino-español que hemos seleccionado. Así que, por parte del docente, antes de empezar, es recomendable que el

profesor haga una breve introducción y explicación sobre los corpus paralelos, más especialmente, estos cinco corpus. Además, el profesor puede hacer una búsqueda de una palabra determinada para ejemplificar el uso. Por parte de los alumnos, ellos pueden hacer la búsqueda a partir de sus propias necesidades e interés, compartir las locuciones que les resulten más interesantes y escoger cualquier locución que les interese para redactar un texto, de ahí que se aumente la autonomía del aprendizaje de la lengua meta, como hemos mencionado en el apartado 3.1.4. En el apartado 5.3, hemos visto que el corpus *Eshelper* tiene la particularidad de presentar las unidades fraseológicas en la página de resultados. Así que el profesor debe pedir a los alumnos que, primero, busquen en este corpus para encontrar locuciones que incluyan la palabra buscada y, a continuación, busquen en otros corpus para obtener más ejemplos. Vemos, por ejemplo, que si un alumno hace una búsqueda de la palabra *mano* en *Eshelper*, puede obtener los resultados que se muestran en la Ilustración 19. A partir de ahí, el alumno puede escoger la locución *de primera mano* y buscarla de nuevo en *ReversoContext* (Ilustración 20) para conseguir más ejemplos de uso.

The screenshot shows the search results for the word "mano" in the Eshelper corpus. The interface includes a search bar with "mano" entered, and several tabs for different types of results: "西英-汉词典" (English-Spanish Dictionary), "西语动词变位" (Spanish Verb Conjugation), "西语例句库" (Spanish Example Sentences), "近反义词" (Near and Far Synonyms), "西西词典" (Spanish-Spanish Dictionary), "西语维基词典" (Spanish Wikipedia Dictionary), "西英词典" (English-Spanish Dictionary), and "全文检索" (Full Text Search). The "西英词典" tab is selected, showing the following results:

- de ~ en ~**
 1. 手递手地 (传递):
Los cubos de agua pasaban de ~ en ~ para apagar el incendio. 人们传水救火.
 2. 一代一代地:
Esta tradición ha llegado de ~ en ~ hasta nosotros. 这个传统一代一代地传给了我们.
- de ~s a boca**
突然地, 突如其来地:
Al cruzar la esquina me lo encontré de ~s a boca. 我在拐过街角时碰上了他.
- de mí < tu > ~**
出于我<你>之手:
Ese cuadro no es de la misma ~ que este otro. 那幅画和这幅画不是同一人的作品.
- de primera ~**
«comprar, adquirir, tomar; saber» 第一手的, 直接 (得到) 的:
Los dirigentes tienen que adentrarse en masas para adquirir informaciones de primera ~. 领导必须深入群众, 以便得到第一手材料.
- descargar la ~**
参见 asentar la ~.

Ilustración 19 Búsqueda de la palabra *mano* en *Eshelper*

The screenshot shows the Reverso Context website interface. At the top, there's a navigation bar with links: TRADUCCIÓN | DICcionario | CORRECTOR | CONJUGACIÓN | SINÓNIMOS | DOCUMENTS. The main search area has a text input with 'de primera mano', a language dropdown set to 'Español', and another dropdown set to 'Chino' with a 'BETA' label. Below the search bar, it says 'Traducción de "de primera mano" en chino'. There are tabs for 'Adjetivo' and 'Adverbio'. Under 'Adjetivo', there are buttons for '第一手', '直接', '亲自', and '睹'. The results section shows five examples of the phrase in context, with the Spanish text on the left and the Chinese translation on the right. The Chinese text is highlighted in yellow.

Spanish Text	Chinese Translation
Lo he visto de primera mano.	我已经看到了第一手的。
La delegación pudo hacer acopio de información de primera mano y entrevistar a testigos presenciales.	该代表团得以收集到第一手资料并约谈证人。
Recibe información de primera mano y puede captar la atmósfera de la situación.	她接收第一手资料，能掌握有关实际情况。
Considera que su informe detallado, de primera mano y objetivo describe claramente la situación en el terreno.	我们认为他们提出的详细和客观的第一手报告明确阐述了该地区的实际情况。
El Comité y el Grupo de Vigilancia deben contar con información de primera mano.	委员会和监测小组应该掌握第一手资料。

Ilustración 20 Búsqueda de la combinación léxica *de primera mano* en *ReversoContext*

Por último, en la actividad 5, la consulta autónoma (de tipo *self-access*) iniciada por los alumnos, como se ha visto en el apartado 3.2.4.3, favorecerá la expresión escrita, puesto que la observación repetida de las concordancias paralelas les ayudará a entender el significado y el uso de determinadas estructuras lingüísticas de manera inductiva (Cruz, 2017). Así pues, en estas cinco actividades, hemos visto cómo usar directamente los corpus paralelos en el aula para aprender locuciones y cómo usarlos como herramientas de consulta para mejorar la expresión escrita. Conviene destacar que, en estas cinco actividades, los corpus paralelos han mostrado su ventaja de ayudar a los alumnos a establecer la correspondencia entre dos lenguas. A continuación, veremos otra propuesta diseñada a partir de la contrastividad entre chino y español.

6.3 Actividades para aprender marcadores del discurso

Otro ejemplo del uso de los corpus paralelos chino-español que proponemos en este trabajo sería para el desarrollo de la competencia discursiva. Más concretamente, vamos a centrarnos en el aprendizaje de los conectores concesivos, especialmente en el conector *aunque*. No obstante, conviene aclarar que, igual que en las actividades para aprender las locuciones, considerando los objetivos de este trabajo, no vamos a adentrarnos en la reflexión teórica sobre los marcadores del discurso⁵⁷ sino centrarnos en las aplicaciones didácticas de los corpus paralelos. Desde nuestro punto de vista, los corpus paralelos podrían ser de gran utilidad en la enseñanza de los marcadores discursivos, puesto que en un diccionario convencional no se puede encontrar una “definición” de un marcador y quizás la mejor manera de saber qué significa un marcador es verlo en su uso y contexto. Además, tras observar las concordancias paralelas, los aprendices pueden comparar las similitudes y diferencias del uso de los marcadores en dos lenguas, extraer su propia conclusión de modo inductivo y reflexionar sobre sus errores interlingüísticos causados por las interferencias de su lengua materna, como hemos dicho en el apartado 3.2.4. A continuación, presentamos nuestra segunda propuesta, en este caso para trabajar los conectores concesivos, especialmente el conector *aunque*, dirigida a los alumnos sinohablantes de nivel intermedio. Los objetivos que pretendemos alcanzar con estas actividades consisten en que los alumnos sean conscientes de que el conector concesivo *aunque* en español puede corresponder a varios conectores en chino (*p. ej.* 虽然, 即使, 尽管, 纵使, etc.), comparen las similitudes y diferencias de su uso entre chino y español, investiguen los diferentes usos del modo verbal en las construcciones con *aunque* y sean conscientes de las diferentes implicaciones que comportan según distintos contextos y traducciones.

⁵⁷ Sobre los marcadores del discurso en chino y español, puede verse Yao (2008); Wang (2013); Chen (2015); Jia (2018).

ACTIVIDAD 1

Aunque existe un diferencial salarial, está disminuyendo.

虽然存在一定的工资差距，但这一差距正在缩小。

(Fuente: *TAUS Data Cloud* y elaboración propia)

1. En este ejemplo arriba, has visto que para expresar 虽然 en español, se usa el conector *aunque*. Busca en *TAUS Data Cloud*, *Glosbe*, *ReversoContext*, *Eshelper* y *CPEIC* el conector *aunque* y observa los resultados. ¿Has encontrado otra traducción de este conector además de 虽然? y, ¿has observado alguna similitud o diferencia del uso de estos conectores en chino y en español?

Al igual que en la primera propuesta, empezamos aquí la actividad 1 con un ejemplo extraído de un corpus paralelo (en este caso, el corpus *TAUS Data Cloud*) para introducir el tema. En este ejemplo, el conector ha sido traducido a 虽然. Sin embargo, el conector *aunque* puede corresponder a muchos marcadores en chino en diferentes contextos (p. ej. 虽然, 即使, 尽管, 纵使, etc.). A modo de ejemplo, si un alumno hace la búsqueda de la palabra *aunque* en el corpus *ReversoContext*, va a obtener los resultados que se muestran en la ilustración 21:

Reverso Context new TRADUCCIÓN | DICcionario | CORRECTOR | CONJUGACIÓN | SINÓNIMOS | DOCUMENTS

aunque Español Chino BETA

Traducción de "aunque" en chino

Adverbio Otro

虽然 尽管 即使 委 甚 纵然 算 怕 承 许 赞 尚 前 特

Ha habido algunos documentos notables, aunque no muchos. 解密内容包括某些著名文件, 虽然数量不算很多。

Hemos obtenido varios otros logros, aunque han sido modestos. 我们还取得了若干其它成就, 虽然也许是微不足道的。

Los progresos han sido innegables, aunque desiguales. 取得的进展是无可否认的, 尽管还不均衡。

Pero debes saberlo, aunque te duela. 但是你应该知道, 尽管这样会伤害你。

No aunque haya una guerra en tu umbral. 没有人能阻止爱情, 即使你面临着战争也

Ilustración 21 Búsqueda de la palabra *aunque* en ReversoContext

Igual que en el estudio de Barlow (2000), tras buscar una determinada palabra y observar las concordancias paralelas repetidamente, los alumnos pueden ser conscientes de que una palabra puede corresponder a varias unidades léxicas en distintos contextos. Además, esta actividad 1, cuya duración es de unos 20 minutos, también tiene el objetivo de que los alumnos descubran las similitudes y diferencias entre el chino y el español en cuanto al uso de estos marcadores. Por ejemplo, a partir del ejemplo que se ha mostrado más arriba, en la actividad 1 (*Aunque existe un diferencial salarial, está disminuyendo. 虽然存在一定的工资差距, 但这一差距正在缩小。*), se ha visto que el fragmento tanto en español como en chino ha sido dividido en dos unidades y las primeras unidades en dos lenguas empiezan con un marcador del discurso que muestra una relación concesiva. Por otra parte, el uso de los marcadores discursivos ha mostrado cierta contrastividad en dos lenguas. De acuerdo con Cao, da

Cunha y Iruskieta (2016), para expresar el mismo significado, es necesario incluir dos marcadores del discurso en este enunciado en chino: 虽然 (*aunque*) en el principio de la primera unidad y 但 (*pero*) en el principio de la segunda unidad. No obstante, en español solo se usa un marcador (*aunque*) en el principio de la primera unidad. Tras comparar repetidamente las concordancias paralelas en dos lenguas, esperamos que los alumnos investiguen este tipo de contrastividad ellos mismos y consigan evitar este tipo de error interlingüístico causado por la interferencia de su lengua materna en su producción. Así, en esta actividad 1, los alumnos se han convertido en investigadores, lo cual aumentará la autonomía del aprendizaje. Conviene señalar que, al igual que en la primera propuesta, antes de empezar, el profesor debe hacer una breve introducción de los corpus paralelos y mencionar que la calidad de los ejemplos no es siempre alta como hemos dicho en el apartado 5.6, pensando en que en las siguientes actividades los alumnos tienen que compilar ejemplos para elaborar sus propias hipótesis. Por eso los alumnos tienen que tener mucho cuidado con los enunciados que no están bien redactados

ACTIVIDAD 2. Observa los ejemplos y presta especial atención al modo del verbo que introduce el conector *aunque* en las oraciones concesivas. ¿Siempre se utiliza el modo indicativo?

ACTIVIDAD 3. Compila 10 ejemplos respectivamente que incluyen las construcciones *aunque* + *indicativo* y *aunque* + *subjuntivo*, e intenta explicar la diferencia según el contexto y la traducción.

Las actividades 2 y 3, que duran 20 minutos aproximadamente, tienen el objetivo de hacer consciente al alumno de que el empleo de los distintos modos verbales⁵⁸ en las construcciones con *aunque* puede comportar diferentes implicaciones. En la actividad 2, los alumnos pueden hacer búsqueda ellos mismos y encontrar las construcciones: *aunque* + *indicativo* y *aunque* + *subjuntivo*. A modo de ejemplo, mostramos la búsqueda de *aunque* en el corpus *Eshelper*.

aunque 专八

音标: [ˈawɲke] 发音: 全球发音: 生词本: 添加笔记:

西汉-汉西词典 西语例句库 近反义词 西西词典 西语维基词典 西英词典 全文检索

《词汇分类教学》 有奖纠错 | 划词

4. Bien, **aunque** todavía no la he visto. 不错吧, 虽然我还没看到她。《NUEVO ELE INICIAL 2》
5. **Aunque** quisiera, no podría seguir enfadado contigo. 即便我想记得, 也不会对你生气的。《不良教育》
6. **Aunque** no lo vayamos a volver a ver. 即使我们不会再相见了。《萌指大叔教你西语对话》
7. Prueba a enviarle un currículum a Satanas, **aunque**... 尝试发一封邮件给撒旦, 尽管... 《影视资讯精选》
8. Sí. **Aunque** la dosificación fuera un bastante alta. 嗯, 虽然药的剂量很高, 《名侦探柯南》
9. ¡**Aunque** todavía no se me dan muy bien! 尽管我弹得还不好! 《萌指大叔教你西语词汇》

Ilustración 22 Búsqueda de la palabra *aunque* en *Eshelper*

Una vez obtenidos suficientes ejemplos, los alumnos pueden observar los contextos y las traducciones de cada construcción y elaborar sus propias hipótesis sobre las implicaciones que se derivan del uso de los diferentes modos verbales. A continuación, vamos a tomar como ejemplo los resultados encontrados en *ReversoContext* para ver qué hipótesis pueden tener los alumnos observando las concordancias paralelas.

⁵⁸ En esta propuesta dejamos al margen los diferentes tiempos verbales y solo centramos en los modos verbales en las construcciones con *aunque*.

Conviene señalar que quizás el corpus más adecuado para esta actividad sea *ReversoContext* porque este corpus tiene una función particular de recuperar el contexto más amplio de cada ejemplo, como se ha expuesto en el apartado 5.6, lo cual puede proporcionar más información de cada situación y facilitar la comparación de cada ejemplo. En la página de la consulta de la palabra *aunque* en *ReversoContext*, hemos escogido 20 ejemplos aleatoriamente que incluyen la construcción *aunque* + *indicativo* y *aunque* + *subjuntivo* (10 ejemplos para cada construcción). Se observa que el conector *aunque* puede corresponder a diferentes conectores en chino según distintos contextos. En cuanto a la construcción *aunque* + *indicativo*, *aunque* ha sido traducido como 虽然 en 5 ejemplos, 尽管 en 4 ejemplos y 不过 en un ejemplo. Mientras que en la construcción *aunque* + *subjuntivo*, *aunque* ha sido traducido como 即使/即便 en 6 ejemplos, 就算 en 2 ejemplos y 尽管 en 2 ejemplos (Tabla 2).

	Aunque + indicativo	Aunque + subjuntivo
虽然	5 ejemplos	0
尽管	4 ejemplos	2 ejemplos
不过	1 ejemplo	0
即使	0	3 ejemplos
即便	0	3 ejemplos
就算	0	2 ejemplos
En total	10 ejemplos	10 ejemplos

Tabla 2 Traducción de *aunque* en 20 ejemplos. Elaboración propia.

A partir de la tabla 2, se observa que los diferentes modos verbales en la construcción con *aunque* influyen su traducción en chino. Por ejemplo, solo en la construcción con

el subjuntivo el conector *aunque* puede corresponder a 即使/即便/就算⁵⁹, los cuales muestran los matices semánticos con otros conectores en chino (虽然, 尽管, 不过). De ahí que los alumnos puedan extraer la conclusión de que los diferentes modos verbales en la construcción con *aunque* pueden tener distintas implicaciones. Gracias a los corpus paralelos, los alumnos tienen la oportunidad de observar la contrastividad entre chino y español, familiarizarse con las estructuras que incluyen *aunque* y las posibles expresiones en chino y comparar el uso de este conector en distintos contextos con la ayuda de la traducción y el contexto.

6.4 Otras posibilidades

En las dos propuestas presentadas en 6.2 y 6.3, hemos explorado la utilización de los corpus paralelos en la enseñanza del ELE para el desarrollo de la competencia léxica y discursiva. Se ha visto que, en comparación con los corpus monolingües, los corpus paralelos tienen la ventaja de mostrar la correspondencia y contrastividad entre dos lenguas, como hemos dicho en el apartado 3.2.4. Por lo tanto, con la finalidad de utilizar esta herramienta especialmente para los alumnos sinohablantes de español, sería recomendable prestar especial atención a los análisis contrastivos entre chino y español y tomarlos como punto de partida a la hora de diseñar las actividades. Aparte de los dos ejemplos que hemos propuesto, los corpus paralelos también pueden aportar muchas posibilidades de la enseñanza del ELE a nivel sintáctico. Así, para aprender las estructuras de la voz pasiva en español, los alumnos pueden hacer búsqueda de la palabra “被”⁶⁰ u otros marcadores de la voz pasiva en chino con el objetivo de obtener ejemplos en voz pasiva en español y familiarizarse con las estructuras. Además, para el

⁵⁹ Según Xu (2016), se usa 即使 para expresar una relación concesiva hipotética. Es decir, la condición que pone son los hechos todavía no realizados o los hechos que no son verdades.

⁶⁰ Un marcador de la voz pasiva en chino.

aprendizaje de los refranes, sería muy beneficioso si los alumnos pudieran compilar sus propios corpus de refranes recogiendo datos e información desde los corpus paralelos. Como señala Valverde (2012), tener la posibilidad de construir su propio corpus supone un ejercicio muy beneficioso para el aprendiente.

Sin duda, son variados los usos de los corpus paralelos en la enseñanza de lenguas. Y es que, gracias a los datos auténticos que nos proporcionan, los alumnos tienen la oportunidad de explorar, aprender e investigar ellos mismos con mayor autonomía.

7. CONCLUSIONES Y LÍNEAS FUTURAS DE INVESTIGACIÓN

El presente trabajo ha explorado la aplicación didáctica de los corpus paralelos chino-español en la enseñanza del ELE con especial atención a los estudiantes sinohablantes. Para poder lanzar nuestra propuesta de actividades, se ha revisado la bibliografía sobre la LC, los corpus paralelos y sus usos prácticos en el ámbito del ELE (objetivo 1). Después, se han establecido unos criterios para seleccionar los corpus adecuados para utilizar en el aula (objetivo 2). Por último, hemos mostrado unas propuestas de actividades y posibilidades utilizando los corpus paralelos (objetivo 3).

Al principio de este trabajo hemos planteado dos preguntas. En la primera pregunta, nos planteábamos qué posibilidades pueden aportar los corpus paralelos en la enseñanza del ELE, especialmente para los alumnos sinohablantes. En el apartado 3, hemos visto que los estudios distinguen cuatro tipos de usos de los corpus paralelos en este ámbito, que son los siguientes: uso directo de las concordancias paralelas en el aula, uso indirecto de los corpus paralelos para la elaboración de materiales didácticos, uso como herramienta de consulta para mejorar la expresión escrita y uso como una base de la investigación en lingüística contrastiva. En cuanto a la segunda, nos preguntábamos entre los corpus paralelos chino-español disponibles en línea actualmente cuáles son los adecuados para emplear en el aula de ELE. Creemos que en el apartado de metodología y presentación de los corpus hemos visto que se pueden utilizar en el aula de ELE los cinco corpus: *TAUS Data Cloud*, *Glosbe*, *Eshelper*, *ReversoContext* y *CPEIC*, y cada uno de ellos nos ha mostrado sus propias ventajas y deficiencias.

Respecto al marco teórico, en primer lugar, se ha visto qué es un corpus y los conceptos relevantes (concordancias, listados de frecuencia y lematización) a la hora de investigar en este ámbito. Con respecto a la aplicación de los corpus en la enseñanza de lenguas, nos ha llamado especialmente la atención el uso directo de los corpus. Después, se ha presentado más concretamente qué es un corpus paralelo y cómo se construye con el objetivo de poder entender mejor los cuatro tipos de usos prácticos que aporta, con especial atención a los usos directos. Cabe señalar que, en nuestro caso, también hemos presentado qué es el “corpuscionario” y la web como corpus, puesto que la mayoría de los corpus estudiados en este trabajo pertenecen a este grupo.

En segundo lugar, en el apartado de metodología y presentación de los corpus, hemos establecido una serie de criterios para seleccionar y comparar los cinco corpus paralelos chino-español disponibles en línea: *TAUS Data Cloud*, *Glosbe*, *Eshelper*, *ReversoContext* y *CPEIC*. Asimismo, se han observado algunas deficiencias, como la baja calidad de texto, la falta de lematización e información anotada, etc. Por lo tanto, consideramos oportuno utilizar los cinco corpus seleccionados de forma combinada para obtener más ejemplos, teniendo en cuenta el nivel de los alumnos y la instrucción del docente.

En tercer lugar, hemos presentado dos propuestas de actividades utilizando los cinco corpus seleccionados para aprender locuciones y conectores respectivamente y hemos mencionado otras posibilidades para el desarrollo de diferentes competencias. Hemos observado que, con la ayuda de los corpus paralelos, los alumnos tienen la oportunidad de establecer correspondencias e investigar la contrastividad ellos mismos. Además, los abundantes contextos y traducciones que ofrecen los corpus paralelos sirven de gran ayuda para trabajar numerosas dificultades en el aprendizaje de la lengua meta.

En conclusión, este trabajo de máster nos ha mostrado múltiples posibilidades del uso de los corpus paralelos en el aula de ELE y ha constituido un intento de ilustrar los posibles usos didácticos de esta herramienta especialmente para los estudiantes sinohablantes. Además, con la ayuda de las descripciones de cada corpus paralelo seleccionado, se han reflexionado sobre los problemas y deficiencias que tienen los corpus paralelos chino-español disponibles en línea hasta la fecha. Esperamos que este trabajo puede servir de ayuda tanto para los investigadores como para cualquier persona (profesor o estudiante) que esté interesada en utilizar los corpus paralelos en el aula.

A pesar todo, este trabajo presenta ciertas limitaciones y muchos aspectos que se pueden mejorar y ampliar. Por ejemplo, considerando la extensión de este TFM, nos limitamos a lanzar unas ideas con ejemplos de actividades. Una posible mejora puede ser diseñar unas unidades didácticas completas a partir de los análisis contrastivos entre chino y español con el objetivo de aprovechar la principal ventaja que aportan los corpus paralelos. Además, también puede ser interesante explorar el uso didáctico del otro tipo de corpus multilingüe, es decir, los corpus comparables, los cuales tienen su propia particularidad y pueden ofrecer otras posibilidades en el aprendizaje del ELE, o utilizar los corpus monolingües y multilingües de forma combinada en el aula. Asimismo, como hemos dicho, la mayoría de los corpus paralelos chino-español disponibles en línea tienen la deficiencia de la baja calidad, así que sería plausible abordar un estudio para construir un corpus paralelo chino-español dirigido por expertos, como el proyecto PaChEs, pero con más variedades de español, a fin de ofrecer un recurso lingüístico representativo y multifuncional para satisfacer las necesidades de diferentes usuarios, especialmente de los alumnos sinohablantes.

8. BIBLIOGRAFÍA

- Alonso Jiménez, E. (2013). Linguee y las nuevas formas de traducir. *Skopos: revista internacional de traducción e interpretación* 2, 5–28. https://rio.upo.es/xmlui/bitstream/handle/10433/851/2013_Alonso_Linguee_Skopos.pdf?sequence=1
- Alonso Pérez-Ávila, E. (2006). El corpus lingüístico en la didáctica del léxico en el aula de E/LE. *RedELE. Red Electrónica de Didáctica del Español como Lengua Extranjera*, 11, 1-83. <https://www.educacionyfp.gob.es/dam/jcr:c9948929-a65a-4afc-8df2-9c104e1e6175/2010-bv-11-01alonso-pdf.pdf>
- Alonso Ramos, M. (2009). Hacia un nuevo recurso léxico: ¿fusión entre corpus y diccionario? *A Survey of Corpus-based Research. Panorama de investigaciones basadas en corpus*, Murcia: Aelinco, 1191-1207. <http://www.dicesp.com/app/webroot/files/file/CILC%2009.pdf>
- Banchs, R. E., Crego, J. M., Lambert, P. y Mariño, J. B. (2006). A Feasibility Study For Chinese-Spanish Statistical Machine Translation. *Proc. of the 5th Int. Symposium on Chinese Spoken Language Processing (ISCSLP)*, Singapore: Springer, 681-692. https://isca-speech.org/archive_open/archive_papers/iscslp2006/B67.pdf
- Barlow, M. (2000). Parallel Texts in Language Teaching. *Multilingual Corpora in Teaching and Research*. Amsterdam: Rodopi, 106–115. https://books.google.es/books?id=eBwAj1nyRu4C&printsec=frontcover&hl=es&source=gbs_ge_summary_r&cad=0#v=onepage&q&f=false
- Bluemel, B. (2014). Learning in Parallel: Using Parallel Corpora to Enhance Written Language Acquisition at the Beginning Level. *Dimension*, 31, 48. <https://files.eric.ed.gov/fulltext/EJ1080263.pdf>
- Braune, F. y Fraser, A. (2010). Improved unsupervised sentence alignment for symmetrical and asymmetrical parallel corpora. *COLING '10 Proceedings of the*

- 23rd International Conference on Computational Linguistics: Posters, Stroudsburg, PA: Association for Computational Linguistics, 81-89. <https://pdfs.semanticscholar.org/4a70/b74711b5a42eedb3a7a785e1c962e2ab56b3.pdf>
- Buyse, K. (2011). ¿Qué corpus en línea utilizar para qué fines en la clase de ele? *Del texto a la lengua: la aplicación de los textos a la enseñanza-aprendizaje del español L2-LE*, Salamanca: Asociación para la Enseñanza del Español como Lengua Extranjera, 277-288. https://cvc.cervantes.es/ensenanza/biblioteca_ele/asele/pdf/21/21_0277.pdf
- Buyse, K. y Verlinde, S. (2013). Possible Effects of Free on Line Data Driven Lexicographic Instruments on Foreign Language Learning: The Case of Linguee and the Interactive Language Toolbox. *Procedia - Social and Behavioral Sciences*, 95, 507-512. <https://doi.org/10.1016/j.sbspro.2013.10.675>
- Buyse, K. (2014). Una hoja de ruta para integrar las TIC en el desarrollo de la expresión escrita recursos y resultados. *Journal of Spanish Language Teaching*, 1(1), 101-115. <https://doi.org/10.1080/23247797.2014.898516>
- Buyse, K. (2021, *en prensa*): Corpus textuales de nativos para investigar sobre la enseñanza/aprendizaje del español LE/L2. *e-Research y español LE/L2. Investigar en la era digital*, Chapt. 7. (Routledge Advances in Spanish Language Teaching). London, New York: Routledge. <https://www.routledge.com/e-Research-y-espanol-LEL2-Investigar-en-la-era-digital/Pinol/p/book/9781138359741>
- Calero, M. Á., Serrano Zapata, M. y Gómez-Devís, M. B. (2020). Codificación y etiquetado en los corpus de aprendices y su aplicación didáctica: la propuesta del corpus de interlengua española de aprendices sinohablantes (cineas). *Revista digital de lingüística aplicada*, 6, 206-222. <https://repositori.udl.cat/bitstream/handle/10459.1/68354/029863.pdf?sequence=1>

- Cao, S. (2018). *Using annotated discourse information of a RST Spanish-Chinese treebank for translation and language learning tasks*. Universitat Pompeu Fabra. <https://www.tdx.cat/bitstream/handle/10803/664419/tsc.pdf?sequence=1>
- Cao, S., da Cunha, I. y Iruskieta, M. (2016). A Corpus-based Approach for Spanish-Chinese Language Learning. *Proceedings of the 3rd Workshop on Natural Language Processing Techniques for Educational Applications (NLPTEA2016)*, Osaka: The COLING 2016 Organizing Committee, 97-106. <https://www.aclweb.org/anthology/W16-4913/>
- Cao, S., da Cunha, I. y Iruskieta, M. (2018). The RST Spanish-Chinese Treebank. *Proceedings of the Joint Workshop on Linguistic Annotation, Multiword Expressions and Constructions (LAW-MWE-CxG-2018)*, Santa Fe: Association for Computational Linguistics, 156-166. <http://ixa2.si.ehu.es/diskurtsoa/rstfilo/>
- Castillo Rodríguez, C. (2009). La elaboración de un corpus ad hoc paralelo bilingüe. *Tradumàtica*, 07. http://www.lexytrad.es/assets/castillo2009a_0.pdf
- Centelles, J. y Costa-Jussà, M. R. (2014). Chinese-to-spanish rule-based machine translation system. *Proceedings of the 3rd Workshop on Hybrid Approaches to Machine Translation (HyTra)*, Gothenburg: Association for Computational Linguistics, 82-86. <https://www.aclweb.org/anthology/W14-1015.pdf>
- Chang, B. (2004). Chinese-English Parallel Corpus Construction and its Application. *Proceedings of the 18th Pacific Asia Conference on Language, Information and Computation*, Waseda University: Logico-Linguistic Society of Japan, 283-290. <http://aclweb.org/anthology/Y04-1030>
- Chen, Y.-L. (2015). *Los marcadores discursivos del chino y del español en textos escritos: los conectores contraargumentativos*. Universidad Autónoma de Madrid. https://repositorio.uam.es/bitstream/handle/10486/672364/chen_ya_ling.pdf?sequence=1
- Corpas Pastor, G. (1996). *Manual de fraseología española*. Madrid: Gredos.

- Costa-Jussà, M. R., Henríquez, C. A. y Banchs, R. E. (2012). Evaluating Indirect Strategies for Chinese–Spanish Statistical Machine Translation. *Journal of Artificial Intelligence Research*, 45, 761-780. <https://arxiv.org/pdf/1402.0563.pdf>
- CPEIC (2017). *Introducción de Corpus Paralelo de Español, Inglés y Chino, CPEIC Corpus*. http://140.116.245.146/cpeic/documents/CP-Introduccion_esp_80213F.pdf
- Cruz, M. (2017). *Lingüística de corpus y enseñanza del español como 2/L (segunda edición)*, Madrid: Arco/Libros (2ª edición).
- Cruz, M. (2015). Léxico y ELE: enseñanza/aprendizaje con tecnologías. *Journal of Spanish Language Teaching (JSLT)*, 2.2, 1-15. <http://diposit.ub.edu/dspace/handle/2445/98868%0A>
- Doval, I. (2017). La construcción de un corpus paralelo bilingüe multifuncional. *Moenia*, 27, 125-141. <https://revistas.usc.es/index.php/moenia/article/view/4071>
- Doval, I. (2018). Corpus paralelos en la enseñanza de lenguas extranjeras: un ejemplo de aplicación basado en el corpus PaGeS. *CLINA: Revista Interdisciplinaria de Traducción, Interpretación y Comunicación Intercultural*, 4(2), 65. <https://doi.org/10.14201/clina2018426582>
- Doval, I., Orduña, J. y Wang, X. (2019). *s.a. Protocolo para constituir el corpus chino-español del Proyecto PaChEs. [Santiago de Compostela / Barcelona; Manuscrito]*.
- Egido, M. y Meliss, M. (2017). ¿Cómo usar diccionarios bilingües en línea? Ejemplos para el desarrollo de las competencias lexicográficas en la enseñanza. *Revista Internacional de Lenguas Extranjeras*, 07, 31-53. <https://revistes.urv.cat/index.php/rile/article/view/1911>
- Flores Acuña, E. (2001). Los diccionarios bilingües como herramientas de referencia para estudiantes avanzados de español como lengua extranjera. *Lenguas modernas*, 29(28), 191-203. <https://revistas.uchile.cl/index.php/LM/article/view/45438>

- Frankenberg-Garcia, A. (2005). Pedagogical uses of monolingual and parallel concordances. *ELT Journal*, 59(3), 189-198. <https://doi.org/10.1093/elt/cci038>
- Frankenberg-Garcia, A. (2004). Lost in parallel concordances. *Corpora and language learners*, Amsterdam: John Benjamins, 213-229. https://www.researchgate.net/publication/253553681_Lost_in_parallel_concordances
- Gelbukh, A., Sidorov, G. y Chanona-Hernández, L. (2002). Corpus virtual, virtual: Un diccionario grande de contextos de palabras españolas compilado a través de Internet. *Proc. Multilingual Information Access and Natural Language Processing, International Workshop, in IBERAMIA-2002, VII Iberoamerican Conference on Artificial Intelligence*, Sevilla: Iberamia, Elsnets y Ritos, 7-14. https://www.academia.edu/3090231/Corpus_virtual_virtual_Un_diccionario_grande_de_contextos_de_palabras_espa%C3%B1olas_compilado_a_trav%C3%A9s_de_Internet
- Hallebeek, J. (1999). El corpus paralelo. *Procesamiento del lenguaje natural*, 24, 58-69. <https://core.ac.uk/download/pdf/16355934.pdf>
- Hernández, H. (2000). El diccionario en la enseñanza de E.L.E. (Diccionarios de español para extranjeros). *¿Qué Español Enseñar? Norma y Variación Lingüísticas en la Enseñanza del Español a Extranjeros*. Zaragoza: Asociación para la Enseñanza del Español como Lengua Extranjera, 93-103. https://cvc.cervantes.es/ensenanza/biblioteca_ele/asele/pdf/11/11_0093.pdf
- Honeyfield, J. (1989). A typology of exercises based on computer generated concordance material. *Guidelines: A Periodical for Classroom Language Teachers*, 11(2), 42-50. https://archive.org/details/ERIC_ED377672
- Jia, J. (2018). *Estudio contrastivo de los conectores causales y consecutivos entre el español y el chino: una propuesta de aplicación didáctica*. Universitat Autònoma de Barcelona. <https://ddd.uab.cat/record/195355>
- Jiménez, J. y Tang, A. (2018). Producción del sistema vocálico del español por hablantes de chino. *Revista de lingüística teórica y aplicada*, 56(1), 13-34.

https://scielo.conicyt.cl/scielo.php?pid=S0718-48832018000100013&script=sci_arttext

Johns, T. (2000). *Tim Johns' data-driven learning Page*.
http://archive.ecml.at/projects/voll/our_resources/graz_2002?ddrivenlrning/whatisddl/resources/tim_ddl_learning_page.htm

Kenning, M.-M. (2010). What are parallel and comparable corpora and how can we use them. *The Routledge handbook of corpus linguistics*, 487-500.
https://www.researchgate.net/publication/265061773_What_are_parallel_and_comparable_corpora_and_how_can_we_use_them

Li, B. y Liu, J. (2008). Mining Chinese-English Parallel Corpora from the Web. *IJCNLP*, 847-852. <http://www.mt-archive.info/IJCNLP-2008-Li-1.pdf>

Lin, T.-J. (2005). *La adquisición y el uso del artículo por alumnos chinos*. Universidad de Alcalá. <https://dialnet.unirioja.es/servlet/tesis?codigo=135605>

Ling, W., Xiang, G., Dyer, C., Black, A. y Trancoso, I. (2013). Microblogs as parallel corpora. *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, Sofia: Association for Computational Linguistics, 176-186. <https://www.aclweb.org/anthology/P13-1018.pdf>

Liu, C., Cruz Silvia, C., Wang, L. y Way, A. (2018). Pivot Machine Translation Using Chinese as Pivot. *China Workshop on Machine Translation*, 74-85.
<http://www.computing.dcu.ie/~away/PUBS/2018/paper44.pdf>

Liu, C.-L. (2015). Locuciones nominales constituidas por nombres y adjetivos de color en chino mandarín y español. *淡江外語論叢*, 26, 89-116.
<https://www.airitilibrary.com/Publication/alDetailedMesh?docid=15627675-201512-201601080009-201601080009-89-116>

Lu, H.-C. (2013). Creation and application of CPEIC, a trilingual parallel corpus. *AACL 2013*.
<https://researchoutput.ncku.edu.tw/en/publications/creation-and-application-of-cpeic-a-trilingual-parallel-corpus>

- Lu, H. -C. y Cheng, A. -C. (2016). Rasgos semánticos a partir de corpus y su aplicación en el aprendizaje de la estructura ‘ser/estar+adjetivo’ en español. *Círculo de Lingüística Aplicada a la Comunicación*, 68(0), 117-137. <https://doi.org/10.5209/clac.54524>
- Lu, H. C., Tsou, W. y Chen, F. (2017). Globalized rhetoric practices on medical writing. *Círculo de lingüística aplicada a la comunicación*, 69, 217-236. <https://revistas.ucm.es/index.php/CLAC/article/view/55320>
- Lu, H. y Lu, W. (2016). 以西語為第三語之受詞代名詞習得研究. National Cheng Kung University. <http://ir.lib.ncku.edu.tw/handle/987654321/175889>
- Lüdeling, A. y Kytö, M. (2008). Corpora and Language Teaching. *Corpus linguistics: an international handbook*, Berlin/New York: Mouton de Gruyter. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.455.9749&rep=rep1&type=pdf>
- Luo, X. (2017). 基于语料库的中文政治文本英西译本的语言特征差异研究. Beijing Foreign Studies University. <http://cdmd.cnki.com.cn/Article/CDMD-10030-1017170679.htm>
- Martínez, R. A. (2018). Diccionario combinatorio para hablantes de portugués: principios metodológicos. *RILEX. Revista sobre investigaciones léxicas*, 1(1), 54-75. <https://doi.org/10.17561/rilex.v1.n1.3>
- Martos Eliche, F. y Contreras Izquierdo, N. M. (2018). El empleo de corpus para el aprendizaje de secuencias formulaicas en ELE/EL2. La frecuencia de uso en el nivel B2 del PCIC. *CHIMERA: Romance Corpora and Linguistic Studies*, 5(1), 1-26. https://www.academia.edu/37586864/El_empleo_de_corpus_para_el_aprendizaje_de_secuencias_formulaicas_en_ELE_EL2._La_frecuencia_de_uso_en_el_nivel_B2_del_PCIC

- McEnery, T. y Hardie, A. (2012). *Corpus Linguistics: Method, theory and practice*. Cambridge University Press. <https://epdf.pub/corpus-linguistics-method-theory-and-practice2f8645efd2df79651f7a90fe3faf67255873.html>
- McEnery, T. y Xiao, R. (2007). Parallel and Comparable Corpora: What are they up to? *Incorporating Corpora*, 18-31. <https://doi.org/10.21832/9781853599873-005>
- McEnery, T. y Xiao, R. (2011). What corpora can offer in language teaching and learning. *Research in Second Language Teaching and Learning* 2, 364-380. https://www.lancaster.ac.uk/fass/projects/corpus/ZJU/xpapers/McEnery_Xiao_teaching.PDF
- Olimpio de Oliveira, M. A. (2015). La adquisición de las unidades fraseológicas mediante un recurso electrónico: el diccionario Linguee. *Linred: lingüística en la Red*, 13, 12. <https://dialnet.unirioja.es/servlet/articulo?codigo=5500691>
- Olímpio de Oliveira, M.A. y Penadés, I. (2016). Linguee como herramienta de enseñanza-aprendizaje de las unidades fraseológicas. *Computerised and corpus-based approaches to phraseology: monolingual and multilingual perspectives*, Geneva: Editions Tradulex, 367-374. <http://www.tradulex.com/varia/Europhras2015.pdf>
- Penadés, I. (1999). *La enseñanza de las unidades fraseológicas*. Madrid: Arco Libros.
- Penas Ibáñez, M. A. y Zhang, X. (2012). *Polisemia funcional de SER y ESTAR en español y en chino. Las locuciones verbales del diccionario DCLEA. Una contribución al estudio del español como lengua 2 (ELE)*. Saarbrücken: Editorial Académica Española.
- Pérez Hernández, M. C. (2002). 2.6 Herramientas básicas de manejo y análisis de corpórea. *Explotación de los corpórea textuales informatizados para la creación de bases de datos terminológicas basadas en el conocimiento*. <http://elies.rediris.es/elies18/26.html>

- Real Academia Española. (2020). *Banco de datos (CREA) [en línea]. Corpus de referencia del español actual*. <http://www.rae.es>
- Real Academia Española. (2001). *Diccionario de la lengua española*. 23a ed. [Edición electrónica]. Madrid: Espasa Calpe.
- Resnik, P., Olsen, M. B. y Diab, M. (1999). The bible as a parallel corpus: Annotating the «book of 2000 tongues». *Language Resources and Evaluation*, 33(1-2), 129-153. <http://users.umiacs.umd.edu/~resnik/parallel/bible.html>
- Resnik, P. y Smith, N. A. (2006). The Web as a Parallel Corpus. *Computational Linguistics*, 29(3), 349-380. <https://doi.org/10.1162/089120103322711578>
- Ruiz Gurilo, L. (2001). *Las locuciones en español actual*. Madrid: Arco Libros.
- Simões, A. (2004). *Parallel corpora word alignment and applications*, Universidade do Minho. <http://ambs.zbr.pt/publications/msc.pdf>
- Sinclair, J. M. (2004). Corpus and text--basic principles. *Developing linguistic corpora: a guide to good practice*. Oxford: Oxbow Books. http://icar.cnrs.fr/ecole_thematique/contaci/documents/Baude/wynne.pdf
- St. John, E. (2001). A case for using a parallel corpus and concordancer for beginners of a foreign language. *Language Learning and Technology*, 5(3), 185-203. https://scholarspace.manoa.hawaii.edu/bitstream/10125/44573/1/05_03_stjohn.pdf
- Sun, Y. (2019). *Estudio sobre la adquisición de los tiempos aspectuales simples del pasado en español por aprendices chinos a partir de distintas perspectivas lingüísticas*. Universitat de Barcelona. https://84.88.27.106/bitstream/handle/10803/667055/YULIANG%20SUN_TESI_S.pdf?sequence=1&isAllowed=y
- Tiedemann, J. (2016). OPUS--Parallel Corpora for Everyone. *Baltic Journal of Modern Computing*, 384. <https://researchportal.helsinki.fi/en/publications/opus-parallel-corpora-for-everyone>

- Valverde, A. M. (2012). *Análisis de errores de aprendientes de francés LE (FLE) basado en corpus orales*. Universidad Autónoma de Madrid. <https://repositorio.uam.es/handle/10486/10147>
- Vargas Sierra, C. (2006). Diseño de un corpus especializado con fines terminográficos: el corpus de la piedra natural. *Debate terminológico*, 2. <http://rua.ua.es/dspace/handle/10045/9426>
- VV.AA. (2008). *Diccionario de términos clave de ELE [en línea]*. https://cvc.cervantes.es/ensenanza/biblioteca_ele/diccio_ele/diccionario/analisiscontrastivo.htm
- Wang, H., Wu, H., Hu, X., Liu, Z., Li, J., Ren, D. y Niu, Z. (2008). The TCH Machine Translation System for IWSLT. *Proc. of the International Workshop on Spoken Language Translation*, Hawaii: Workshop IWSLT, 124–131. https://www.isca-speech.org/archive/iwslt_08/slt8_124.html
- Wang, L. (2001). Exploring parallel concordancing in English and Chinese. *Language Learning and Technology*, 5(3), 174-184. https://scholarspace.manoa.hawaii.edu/bitstream/10125/44572/1/05_03_wang.pdf
- Wang, Y.-C. (2013). *Los marcadores conversacionales en el subtitulado del español al chino análisis de La mala educación y Volver de Pedro Almodóvar*. Universitat Autònoma de Barcelona. https://ddd.uab.cat/pub/tesis/2013/hdl_10803_125655/ycw1de1.pdf
- Wu, F. (2016). Fraseología comparada del español y del chino: su aplicación a la enseñanza en la clase de español como lengua extranjera. *marcoELE. Revista de Didáctica Español Lengua Extranjera*, 22 (ISSN: 1885-2211). <https://marcoele.com/descargas/22/wu-fraseologia.pdf>
- Xu, S. (2016). “尽管”“虽然”与“即使”的用法分析. *现代语文: 下旬. 语言研究*, (011), 67-70. <http://www.cqvip.com/qk/87478x/201611/670760630.html>

- Yao, G. (2018). *La aplicación del corpus en el aprendizaje del vocabulario de español en las universidades chinas: basada en un estudio empírico*. Universidad de Estudios Internacionales de Shanghai. <http://cdmd.cnki.com.cn/Article/CDMD-10271-1018231284.htm>
- Yao, J. (2008). *Estudio comparativo de los marcadores del discurso en español y en chino mandarín a través de diálogos Cinematográficos*. Universidad de Valladolid. <http://www.educacionyfp.gob.es/dam/jcr:809c6d1c-f283-4036-b34a-3c8c5c578514/2009-bv-10-25yao-pdf.pdf>
- Zanettin, F. (1994). Parallel Words: Designing a Bilingual Database for Translation Activities. *Corpora in Language Education and Reserch: A Selection of Papers from Talc94*, 4, 99-111. https://www.academia.edu/3490875/Parallel_Words_Designing_a_Bilingual_Database_for_Translation_Activities
- Zhang, Y., Wu, K., Gao, J. y Vines, P. (2006). Automatic acquisition of chinese-english parallel corpus from the web. *Advances in Information Retrieval: 28th European Conference on IR Research, ECIR 2006*, London: Springer, 420-431. https://www.researchgate.net/publication/221397486_Automatic_Acquisition_of_Chinese-English_Parallel_Corpus_from_the_Web
- Ziemski, M., Junczys-Dowmunt, M. y Pouliquen, B. (2016). *The United Nations Parallel Corpus, Language Resources and Evaluation (LREC'16)*, Portorož, Eslovenia. <https://conferences.unite.un.org/UNCORPUS/es>

9. ANEXOS

Anexo 1: Enlaces de los corpus paralelos chino-español consultados:

Corpus Paralelo de Español Inglés y Chino (CPEIC):

<http://140.116.245.146/cpeic/TriLin.html?role=teacher>

Corpus Paralelo de las Naciones Unidas: <https://conferences.unite.un.org/uncorpus/>

Eshelper: <https://www.esdict.cn/>

Glosbe: <https://es.glosbe.com/es/zh>

MyMemory: <https://mymemory.translated.net/>

OPUS: <http://opus.nlpl.eu/>

ReversoContext: <https://context.reverso.net/traduccion/>

RST Spanish-Chinese Treebank: <http://ixa2.si.ehu.es/rst/zh/index.php>

TAUS Data Cloud: <https://data-app.taus.net/>

The Holy Bible: <http://users.umiacs.umd.edu/~resnik/parallel/bible.html>

Anexo 2: Fuentes de las concordancias paralelas extraídas de los corpus paralelos:

(para mejor comprensión, algunas traducciones extraídas han sido modificadas)

<https://www.esdict.cn/dicts/es/boca> (para la actividad 1 y 2 en el apartado 6.2)

<https://www.esdict.cn/dicts/es/boca%20abierta> (para la actividad 1 en el apartado 6.2)

<https://www.esdict.cn/dicts/es/segunda%20mano> (para la actividad 3 en el apartado 6.2)

<https://www.esdict.cn/dicts/es/pie> (para la actividad 3 en el apartado 6.2)

<https://www.esdict.cn/dicts/es/abrir%20y%20cerrar%20de%20ojos> (para la actividad 3 en el apartado 6.2)

<https://context.reverso.net/traduccion/espanol-chino/quemar+pesta%C3%B1a> (para la actividad 3 en el apartado 6.2)

<https://www.esdict.cn/dicts/es/mano> (para la actividad 3 en el apartado 6.2)

<https://www.esdict.cn/dicts/es/tomando%20el%20pelo> (para la actividad 3 en el apartado 6.2)

<https://data-app.taus.net/> (para la actividad 1 en el apartado 6.3)

<https://context.reverso.net/traduccion/espanol-chino/aunque> (para la actividad 1 en el apartado 6.3)

Anexo 3: Ejemplos extraídos de ReversoContext que incluyen *aunque* para la actividad 3 en la segunda propuesta:

Hemos obtenido varios otros logros, <i>aunque</i> han sido modestos. 我们还取得了若干其它成就， <i>虽然</i> 也许是微不足道的。
No entendí nada, <i>aunque</i> hablaste mucho. <i>虽然</i> 你讲了很多 但我好像不是很清晰
Quince Partes proporcionaron datos, <i>aunque</i> las hipótesis a veces eran diferentes. 15 个缔约方提供了数据， <i>尽管</i> 其中的设想有时不同。
Es evidente que la mentalidad está cambiando, <i>aunque</i> será un proceso largo. 显而易见，人们的心态正在改变， <i>尽管</i> 这将是一个漫长的过程。
Esta cuestión sigue estando muy controvertida, <i>aunque</i> ha sido objeto de considerables debates. <i>虽然</i> 已经对这个问题进行过大量的讨论，但还是有不少争议。
Esa delegación ha tenido algún éxito, <i>aunque</i> las tensiones continúan. 这个代表团取得了某些成功， <i>虽然</i> 紧张局势依然存在。
La mortalidad materna, <i>aunque</i> está disminuyendo, es todavía elevada. 孕产妇死亡率 <i>尽管</i> 在逐渐下降，但是依然很高。
El rey era el presidente de la Alta Corte, <i>aunque</i> legalmente era primus inter pares. 国王被认为是高等法院的领袖， <i>尽管</i> 在法律上他只是同僚中的首席（primus inter pares）。
Esta posibilidad seguía existiendo, <i>aunque</i> también planteaba problemas. 这仍然是可能的， <i>不过</i> 也会引起某些问题。

Este enfoque, **aunque** es conceptualmente sencillo, plantea retos programáticos y financieros.

虽然在概念上简单，这种做法却有其方案和财务报告方面的困难。

Pero debes saberlo, **aunque** te duela.

但是你应该知道，**尽管**这样会伤害你。

Los padres tienen derechos y obligaciones iguales **aunque** el matrimonio se haya disuelto.

即使婚姻关系已经解除，父母仍享有平等的权利和义务。

Dicha capacidad se mantiene **aunque** se disuelva el matrimonio.

即使婚姻解除这种能力仍全部维持。

Esta inmunidad continuará **aunque** hayan dejado de desempeñar misiones para la Autoridad;

有关人员**即使**已不再受雇为管理局执行任务，仍应继续享有此项豁免；

De todos modos, y **aunque** no tenga mucha lógica, se hicieron amigos.

虽然经常吵架，**尽管**不是特别喜欢，她绝望彼此之间成为朋友。

La mujer recibe su parte legal, **aunque** sea rica.

女性可获得其应有的一份遗产，**即便**她已经很富有。

Detesto pelearme contigo **aunque** sea falso.

我讨厌跟你吵架 **就算是**做戏

Yo quisiera vivir aquí **aunque** tuviera novia.

真的吗？**就算**我有女友 我也会想住在这里

No se permitirían viajes excepcionales **aunque** haya situaciones apremiantes.

即便有紧急情况，也不允许进行特准旅行；

Nos gustaría entablar ese debate interactivo, **aunque** se prevea para finales del período de sesiones.

我们将欢迎举行这样一次交互式辩论，**即便**是在本届会议的今后某个阶段举行。